UNIVERSITÉ DE SHERBROOKE
Faculté de génie
Département de génie électrique et de génie informatique

FLCAA : Système de codage parcimonieux et d'analyse perceptuelle des signaux sonores en temps réel

Mémoire de maîtrise
Spécialité : génie électrique

Vincent Tremblay-Boucher

Jury : Denis Gingras, rapporteur
Ramin Pichevar, membre externe
Jean Rouat, directeur

Sherbrooke (Québec) Canada Décembre 2013

Your file Votre référence
ISBN: 978-0-499-00358-4

Our file Notre référence
ISBN: 978-0-499-00358-4

RÉSUMÉ

Ce mémoire débute par un survol de l'état de l'art des méthodes de compositions musicales assistées par ordinateur (MCMAO). À l'aide d'un ensembles critères permettant l'évaluation des méthodes de compositions musicales assistées par ordinateur, on identifie une technique particulièrement prometteuse. Il s'agit d'un compositeur statistique, présenté par Hoffman *et al.* en 2008, utilisant les "mel-frequecy cepstral coefficients" (MFCC), un prétraitement inspiré des techniques en reconnaissance de parole. Toutefois, cette technique présente diverses limitations, comme la qualité de reconstruction des signaux, qui l'empêche d'être utilisée pour composer de la musique utilisable professionnellement. Ainsi, ce mémoire tente de bonifier la méthode de composition musicale assistée par ordinateur de Hoffman *et al.* en remplacant la technique MFCC d'analyse/synthèse du signal par une technique novatrice d'analyse/synthèse des signaux sonores nommée "Fast Locally competitive algorithm for audio" (FLCAA). Celle-ci permet une analyse perceptuelle parcimonieuse, en temps réel, ayant une bonne résolution fréquencielle et une bonne résolution temporelle. De plus le FLCAA permet une reconstruction robuste de bonne qualité également en temps réel. L'analyse est constituée de deux parties distinctes. La première consiste à utiliser un prétraitement inspiré de l'audition pour transformer le signal sonore afin d'obtenir une représentation cochléaire. Concrètement, la transformation d'analyse est accomplie à l'aide d'un filtrage par banc de filtres cochléaires combiné à un mécanisme de fenêtre coulissante. Le banc de filtres utilisé est composé de filtres cochléaires passe-bande à réponse impulsionnelle finie, de type "rounded exponential" (RoExp). La deuxième étape consiste à coder la représentation cochléaire de manière parcimonieuse afin d'augmenter la résolution spatiale et temporelle pour mettre en évidence certaines caractéristiques du signal comme les fréquences fondamentales, l'information contenue dans les basses fréquences et les signaux transitoires. Cela est fait, en intégrant un réseau de neurones (nommé LCA) utilisant les mécanismes d'inhibition latérale et de seuillage. À partir des coefficients de la représentation perceptuelle, il est possible d'effectuer la transformation de synthèse en utilisant une technique de reconstruction novatrice qui est expliqué en détail dans ce mémoire.

Mots-clés : Méthode de composition musicale assistée par ordinateur (MCMAO), analyse perceptuelle, synthèse bio-inspiré, temps réel, signaux sonores, banc de filtres cochléaires, seuillage, codage parcimonieux

REMERCIEMENTS

Merci à Jean Rouat pour son aide et son appui tout au long de mes recherches et expériences. À Stéphane Molotchnikoff pour nous avoir référé au travail de Christopher J. Rozell lors d'une rencontre hebdomadaire du groupe NECOTIS. À Christopher J. Rozell pour le code Matlab du LCA. À Ramin Pichevar pour ses travaux et discussions sur le LCA adapté pour les signaux sonores. Et, merci à Stéphane Loiselle pour le code utilisé pour la génération d'un banc de filtres cochléaires de type RoExp.

TABLE DES MATIÈRES

1 Introduction **1**
- 1.1 Mise en contexte et problématique . . . 1
- 1.2 Définition du projet de recherche . . . 7
- 1.3 Objectifs du projet de recherche . . . 8
- 1.4 Contribution originale . . . 9
- 1.5 Plan du document . . . 9

2 État de l'art **11**
- 2.1 Filtrage par banc de filtres . . . 11
- 2.2 Banc de filtres cochléaires et représentation perceptuelle . . . 12
- 2.3 Algorithme localement compétitif ou "Locally Competitive Algorithm" (LCA) 13
  - 2.3.1 Description détaillée de l'algorithme LCA . . . 17
- 2.4 Application du LCA pour le codage de signaux sonores . . . 19
- 2.5 Causal Local Competitive Algorithm (CLCA) . . . 21
- 2.6 Réflexion sur l'état de l'art . . . 21

3 FLCAA : Filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante et le codage LCA **23**
- 3.1 Modifications apportées au système LCA . . . 24
  - 3.1.1 Fenêtrage . . . 24
  - 3.1.2 Conception des filtres cochléaires utilisés . . . 26
  - 3.1.3 Transformation d'analyse . . . 27
  - 3.1.4 Inhibition latérale . . . 29
  - 3.1.5 Critères d'optimisation du codage . . . 29
  - 3.1.6 Transformation de synthèse . . . 29

4 Expériences et conditions expérimentales **33**
- 4.1 Environnement et paramètres . . . 33
- 4.2 Validation de la fonctionnalité et de la polyvalence du FLCAA . . . 35
- 4.3 Qualité de la reconstruction . . . 35
- 4.4 Durée nécessaire au codage . . . 35
- 4.5 Robustesse . . . 36

5 Résultats et discussions **37**
- 5.1 Validation de la fonctionnalité et de la polyvalence du FLCAA . . . 37
- 5.2 Qualité de reconstruction . . . 40
- 5.3 Durée nécessaire au codage . . . 44
- 5.4 Robustesse . . . 45

6 Conclusion **47**

LISTE DES RÉFÉRENCES **49**

LISTE DES FIGURES

2.1 Architecture du banc de filtres de Vetterli. 13
2.2 Architecture du système LCA. 15
2.3 Fonctions de seuillage. 16
2.4 Stratégie de codage LCA. 18
2.5 Représentation auditive perceptuelle du système LCAASS. 20

3.1 Architecture du réseau de neurones utilisé par le FLCAA. 24
3.2 Architecture du système FLCAA. 25
3.3 Forme schématisée d'un filtre auditif. 26
3.4 Représentation spectrale du banc de filtres. 28

5.1 Comparaison des représentations perceptuelles du signal /d/ /a/. 39
5.2 Comparaison des représentations perceptuelles du saxophone. 40
5.3 Comparaison des représentations perceptuelles du chant d'une femme. . . . 41
5.4 Comparaison des représentations perceptuelles pour une chanson country folk. 43
5.5 Comparaison des représentations perceptuelles pour une chanson electro. . 44
5.6 Évaluation des durées nécessaires aux codages. 45
5.7 Évaluation de la robustesse du FLCAA 46

LISTE DES TABLEAUX

5.1 Comparaison des performances du LCAASS en fonction du nombre d'itérations 38
5.2 Comparaison des performances du LCAASS et du FLCAA. 42

LEXIQUE

Terme technique	Définition
Parcimonie	Codage utilisant un minimum de coefficients
Inhibition latérale	Stratégie de codage bio-inspirée favorisant la parcimonie.
Cochlée	Organe de l'oreille interne permettant la transduction de la pression acoustique à des trains de potentiels d'action.
Filtre cochléaire	Filtre numérique bio-inspiré de la cochlée
Fonction de coût	Fonction utilisé par le système de codage permettant de déterminer la qualité des solutions envisagées afin de diriger les essaies vers l'obtention d'une solution optimale.
Fenêtre d'analyse	Un segment du signal qui sera analysé.
Dictionnaire Surcomplet	Ensemble de bases permettant de représenter le signal de façon surcomplète. C'est-à-dire que le nombre de bases est bien plus grand que la dimension de l'espace de travail.
Matching Pursuit	Technique itérative numérique d'optimisation.
Norme L0	Norme mathématique assurant l'atteinte d'une solution optimale.
Norme L1	Norme mathématique assurant l'atteinte d'une solution localement optimale.

LISTE DES SYMBOLES

Symbole	Définition
$\lfloor . \rfloor$	Correspond à la partie entière d'un nombre réel
$< . >$	Produit scalaire
$\|.\|$	Valeur absolue
$\|\|.\|\|$	Module
$\sum$	Sommation
$*$	Opération de convolution
$<$	Inférieur strict
$\leq$	Inférieur ou égal
$H(.)$	Fonction de transfert d'analyse
$F(.)$	Fonction de transfert de synthèse
R	Facteur de décimation
Φ	Ensembles des filtres représentant le signal/Dictionnaire surcomplet
m	Indice du neurone traité
t	Instant
ϕ_m	Coefficients du filtre associé au neurone m
G	Matrice d'inhibition latérale
T_λ	Fonction de seuillage
$a(t)$	Coefficients parcimonieux résultant du codage LCA, LCAASS ou FLCAA
$a_m(t)$	Coefficients parcimonieux du LCA, LCAASS ou FLCAA pour le neurone m
$u(t)$	Potentiels internes des neurones durant le codage LCA, LCAASS ou FLCAA
$u_m(t)$	Potentiel interne, du neurone m, durant le codage LCA, LCAASS ou FLCAA
$s(t)$	Signal original
$\hat{s}(t)$	Signal reconstruit
b_m	Signal transformé par projection sur le filtre associé au neurone m
C	Fonction de coût
λ	Paramètre de parcimonie
τ	Facteur d'inertie
K	Nombre de filtres
L	Longueur de $s(t)$ en nombre d'échantillons
q	Facteur de décalage de Φ dans le LCAASS
l	Nombre d'échantillons de décalage entre deux fenêtres d'analyse consécutives
$w[j]$	Fenêtres d'analyse originales
$\hat{w}[j]$	Fenêtres d'analyse reconstruites
L_w	Nombre de fenêtres d'analyse
N	Largeur de la fenêtre d'analyse ou nombre de filtres
Δ	Le pas d'échantillonnage exprimé en millisecondes

LISTE DES ACRONYMES

Acronyme	Définition
DSL	Différence spectrale logarithmique
EPQP	Évaluation perceptuelle de la qualité de la parole
EQM	Erreur Quadratique Moyenne
LCA	Locally Competitive Algorithm
CLCA	Causal Locally Competitive Algorithm
MCMAO	Méthode de composition musicale assistée par ordinateur
LCAASS	LCA adapté aux signaux sonores
FLCAA	Fast Locally Competitive Algorithm for audio
UdeS	Université de Sherbrooke
RIF	Réponse impulsionnelle finie
RoExp	Rounded exponential
RSB	Rapport signal à bruit

CHAPITRE 1

Introduction

1.1 Mise en contexte et problématique

Depuis le début des temps, l'homme est fasciné par la musique. On n'a qu'à penser aux rythmes tribaux ancestraux pour comprendre que la musique a toujours été une forme d'expression intrinsèquement propre à l'homme. En effet, la musique existe depuis les temps les plus anciens, sûrement avant même l'époque de ses premières traces historiques. À notre connaissance, il n'existe pas de civilisation qui, tôt ou tard, n'ait développé son propre système musical ou n'en ait adopté un.

Ainsi, on comprend l'importance historique de la musique pour l'homme et la société. Aujourd'hui la musique est toujours aussi importante pour des centaines de millions de gens sur la planète. On n'a qu'a penser aux artistes interprètes, professionnels ou non, à tous les admirateurs qui se " nourrissent " de musique et à ceux qui utilisent la musique pour bonifier leurs médias (film, télévision, radio, etc.) pour comprendre la place qu'occupe la musique aujourd'hui.

La musique est une forme d'art à part entière étant donné le niveau d'abstraction nécessaire à sa création. Depuis le début des temps, la composition musicale n'était possible que par des humains. Depuis l'apparition de la science informatique, au début des années 50, il y a eu plusieurs recherches ayant pour objectif la création d'un compositeur de musique informatisé. Les méthodes de composition musicale assistée par ordinateur(MCMAO) se sont raffinées avec le temps, chacune ayant ses forces et ses faiblesses.

Suite à notre recherche bibliographique, il a été possible de mettre en évidence 6 critères déterminants pour une bonne MCMAO. Cela est particulièrement intéressant pour pouvoir comparer les MCMAO entre elles et mettre en évidence leurs performances relatives.

1) **Créativité** : La capacité de synthèse de nouvelles sonorités et la capacité d'assemblage de manière novatrices.

2) **Adaptabilité** : L'utilisateur n'a pas besoin de générer de nouvelles règles de composition pour modifier le style musical en sortie du système. Ainsi, le système est plus polyvalent, plus facile à utiliser et moins coûteux.

3) **Autonomie** : Capacité de composer avec peu ou pas d'interactions humaines. Cela permet d'éviter les problèmes d'interface et le biais de l'interaction personne-machine.

4) **Qualité** : Capacité de composer de la musique qui est "bonne" à écouter. La musique ne doit pas être déplaisante à écouter à cause de sons aggressants ou d'enchaînements trop secs. Cette notion est partiellement subjective et doit être évaluée, avec ouverture d'esprit, en considérant autant la qualité objective du signal synthètisé que la qualité subjective de la composition, c'est-à-dire la fluidité des enchaînements et des mélodies. Ce critère est crucial pour l'utilisation d'une MCMAO.

5) **Versatilité** : Capacité de composer tous les styles de musique. Actuellement, la plupart des systèmes composent de la musique classique ayant des régles de composition bien définies. Une bonne versatilité permettrait de composer tous les styles de musique tels que le rock, le pop, le dance, le classique, le heavy metal, le punk rock, le ska, le techno, le classique, etc. Ces types de musique ont des structures musicales et des sonorités diverses qui doivent être représentables par un bon système.

6) **Potentiel** : Capacité de composition de séquences musicales de longueurs indéfinies. Comme la création peut être parfois bonne ou mauvaise, comme pour la composition humaine, il faut un modèle capable de générer plus de séquences sonores que nécessaire. Ainsi, on peut conserver les meilleures séquences et rejeter celles qui ne plaisent pas au compositeur.

Il y a deux catégories de techniques pour composer de la musique à l'aide d'ordinateur. Celles-ci se distinguent par leur méthode de génération des règles de composition. En effet, l'approche explicite nécessite qu'un expert connaissant les règles de composition musicale définisse celles utilisées dans la pièce de référence. Toutefois, selon le style de la pièce de référence, les règles de composition peuvent être très complexes et difficilement identifiables. Ainsi, il est possible de composer de la musique respectant des règles de composition préalablement établies. L'approche implicite quant à elle, permet l'apprentissage des règles de composition à partir des données d'entraînement fournies [8]. Cela permet de ne pas avoir à redéfinir les règles de composition si on change de pièce de référence.

Ce fut Mathews qui débuta, dans les années 50, les recherches sur les MCMAO avec un système implicite utilisant un modèle de Markov. Il publia en 1963 l'avancement de ses recherches [21]. Suite aux premiers insuccès des MCMAO implicites, les recherches se tournèrent vers les MCMAO explicites. Dès 1957 [15] on tentait de définir les règles de composition le plus exactement possible pour arriver à coder ces règles afin qu'un ordinateur puisse générer de la musique.

Depuis, plusieurs techniques de composition explicite virent le jour. On pense notamment au MCMAO à base de grammaires qui consiste à identifier les éléments de base d'un type de musique et à les utiliser en tant que grammaire de composition. Ainsi, on utilise plusieurs bases musicales définies par un expert afin de représenter les signaux sonores. Ces bases sont souvent associées à des statistiques afin de pouvoir générer des séquences musicales. McCormack [22] fait un excellent survol de la technique de composition à base de grammaire. Cette méthode ne permet pas d'obtenir de nouveaux sons ou de nouvelles notes, car elle utilise uniquement la grammaire qu'elle possède. Toutefois, P. Sheikholharam [35] contourne cette limitation en combinant la grammaire avec un algorithme génétique pour produire de nouveaux sons. Il est également possible d'optimiser la grammaire en utilisant un processus d'optimisation nommé le processus de Dirichlet hiérarchique (HDP) tel que fait par Liang *et al.* [18]. Ils représentent la grammaire sous forme d'arbre. Cela à l'avantage de correspondre à une forme d'écriture musicale classique et de faciliter le travail de l'utilisateur. Ces MCMAO ont deux limitations majeures. La première est que la grammaire doit être redéfinie si on change d'interprète ou de style de musique, ce qui démontre peu ou pas d'adaptabilité. La deuxième est qu'il n'y a pas de nouvelles sonorités produites, ce qui démontre peu de créativité. On dénote toutefois qu'une MCMAO hybride incluant un algorithme génétique permet une certaine créativité.

Pour permettre d'améliorer la créativité des MCMAO explicites, Anders [1] développa le système de composition à base de contraintes. Celui-ci doit définir plusieurs règles de composition affectant les mêmes paramètres. Ainsi, il utilise plusieurs règles de composition pour représenter, par exemple, la fréquence fondamentale. Ce mélange de règles permet d'obtenir des séquences musicales nouvelles contrairement à la plupart des méthodes explicites. Afin de pouvoir utiliser la MCMAO à base de contraintes, Anders présente un environnement de composition à base de contraintes qui se nomme PWConstraints [1]. Cet environnement permet à l'utilisateur de définir des règles de composition et les contraintes qui lui sont attachées en utilisant le langage de programmation graphique PatchWork.

Les MCMAO explicites sont souvent simples à valider étant donné qu'on peut vérifier l'atteinte de l'objectif à l'aide d'une fonction d'évaluation basée sur des concepts musicaux. Ils sont intuitifs pour des musiciens qui tentent d'écrire de la musique à l'aide d'ordinateur. Toutefois, ils sont rigides étant donné les règles imposées par l'humain. Si on change de style, il faut recommencer l'étape de l'identification des règles d'écriture musicale ou de la définition de la grammaire ce qui demande beaucoup de travail. De plus, E.R. Miranda mentionne que "*La formalisation de la musique par des règles strictes entraîne, la plupart du temps, une perte d'informations informelles faisant partie intégrante de la musique. La*

musique est en partie un ensemble de sentiments, d'expériences et de culture. Il est donc difficile, si possible, d'exprimer explicitement tout ce qui contribue à la composition de la musique de manière explicite." [24]. Voyant ces limitations évidentes, plusieurs chercheurs tentèrent de les contourner avec diverses méthodes implicites.

Chiu *et al.* [8] font un survol de quelques travaux récents dans le domaine des MCMAO. Voici la traduction d'un court extrait :

" Les travaux récents sur les MCMAO tentent de développer les approches implicites de génération de règles de composition. D. Cope (1992) a séparé de la musique en petits segments. Un nouvel objet musical est généré en analysant et en combinant ces petits segments. Y. Marom (1997) a utilisé les chaînes de Markov pour modéliser la mélodie. Au centre de recherche IRCAM, S. Dubnov et coll. (2003) ont construit un modèle pour simuler le style des grands maîtres en utilisant l'analyse syntaxique incrémentale (IP) et les arbres de prédiction de suffixes (PST). Au CMU, B Thom (2001) a proposé un système permettant l'interaction temps réel entre le système générant un solo dans le style du solo joué par l'utilisateur. Ce système modélise le style du soliste en utilisant un algorithme de maximisation de la correspondance pour générer de la musique. M. Farboot (2001) du MIT a présenté une MCMAO qui génère de la musique selon le concept de la peinture."

Il existe encore plusieurs autres types de MCMAO. On pense par exemple aux MCMAO utilisant des réseaux de neurones [9] [17] [7] [26] [14] [27]. Ceux-ci sont simples à utiliser mais sont difficilement contrôlables ce qui nuit à la qualité de la composition. La composition génétique [4] quant à elle permet beaucoup de créativité mais la composition est lente et de mauvaise qualité [27]. Les systèmes combinant la composition génétique avec l'interaction de l'utilisateur [13] [2] ont permis d'obtenir une meilleur qualité de composition mais la lourdeur de l'interface est problématique.

Le processus de synthèse par concaténation des données utilise une base de données pour y extraire de courtes trames de musique considérées comme les notes de la composition. Plusieurs variantes de cette méthode sont présentées, en 2003, par Schwarz [33]. On compte parmi elles : le " Plunderphonics ", la mosaïque musicale (" Musical Mosaicing "), la synthèse de la voie chantée par concaténation (" Concatenative singing voice synthesis "), l'échantillonnage (" Sampling") et la synthèse granulaire (" Granular synthesis "). Schwarz propose aussi un modèle original, le Catterpillar [33], unifiant les différentes approches. Le Catterpillar utilise une large banque de données permettant d'obtenir un maximum d'information afin de représenter le plus justement les séquences sonores que l'on désire obtenir. Les segments sonores sont choisis selon des fonctions d'optimisation basées sur la

ressemblance de sonorité et de contexte. Casey [5], en 2005, écrit sur la nécessité d'avoir beaucoup de données pour composer une musique de qualité ce qui entraîne un ralentissement de la composition : " *Plusieurs aspects permettent le contrôle de la synthèse comme : le changement des caractéristiques, la modification des fonctions de validation et les données fournies au système. La qualité de la synthèse de ces systèmes tend à s'améliorer si le système dispose de plus de données. Toutefois, plus de données impliquent une croissance exponentielle du temps de recherche du système.*" Afin de contourner cette limitation, Casey pose un niveau d'abstraction supplémentaire en créant un vocabulaire à l'aide des sonorités les plus fréquemment trouvées dans la base de données [5]. Il nomme les éléments de base lexèmes sonores afin de faire le lien avec les éléments de base de la parole. Ces lexèmes sonores permettent de représenter une vaste plage de sonorités. Les MCMAO par synthèse de données permettent, avec beaucoup de données, une bonne qualité musicale mais ne permettent pas beaucoup de créativité car ils sont limités par leur base de données.

Parmi toutes les solutions étudiées et évaluées, c'est le système de Hoffman *et al.* [16] qui répond au plus grand nombre de critères de performance énumérées en début de sous-section. Plus précisemment, ils utilisent un pré-traitement pour extraire les "mel frequency cepstral coefficients" (MFCC) [23] [12] qu'ils utilisent pour entraîner leur système. Ils choisissent la représentation MFCC représentant l'enveloppe du "short term power spectrum" ce qui révèle beaucoup d'informations sur le signal analysé. De plus, ils utilisent un modèle stochastique Markovien, le modèle de Markov caché [36], qu'ils initialisent à l'aide d'un processus de Dirichlet hierarchique [37] tel que présenté par l'équation 1.1.

$$\begin{aligned} \beta_0 &\sim GEM(\delta) \\ \beta_i &\sim DP(\gamma, \beta_0) \\ \pi_{i,j}(\alpha, \beta_i); z_{i,t} &\sim \pi_{z_{t-1}} \\ \theta_k \sim H; y_{i,t} &\sim F(\theta_{z_{i,t}} \end{aligned} \tag{1.1}$$

Où GEM correspond à la construction de Sethuraman [34] et DP au processus de Dirichlet. β_0 est le vecteur de probabilités permettant le choix de la chanson i. Le paramètre δ détermine la variance de β_0. Chaque chanson i à un vecteur de probabilités des états β_i, une matrice de probabilités des transitions entre les états π_i, une séquence d'états z_i, une séquence d'observations y_i et une matrice de probabilités d'émission θ des sorties du modèle. Le paramètre γ détermine la variance de $\pi_{i,j}$ par rapport à β_i tandis que

le paramètre α détermine la variance de $F(\theta)$ par rapport à $\pi_{i,j}$. H est l'ensemble des observations possibles.

Malgré un système fort intéressant de modélisation musicale, les travaux de Hoffman n'ont pas reçus beaucoup d'attention parce que les résultats présentés souffraient d'une reconstruction de piètre qualité. Cela, à cause que les MFCC sont partiellement inversibles car ils ne conservent pas l'information reliée à la phase du signal ce qui entraîne une dégradation significative du signal reconstruit. L'information de phase est perdue parce que la transformation par MFCC se base uniquement sur le spectre de puissance et non pas sur le spectre complet du signal. En conséquence, cela limite la qualité de la musique générée par la méthode de composition musicale assistée par ordinateur de Hoffman *et al.*.

Ainsi, afin de pallier à cette limitation et d'obtenir une reconstruction de qualité, nous proposons l'utilisation de la transformation LCA adaptée aux signaux sonores (LCAASS) présenté par Pichevar *et al.* [29]. Cette méthode exhibe un excellent potentiel pour le traitement de signaux sonores car elle permet l'obtention d'une représentation conservant plus d'informations sur le signal analysé que les MFCC. En effet, la représentation LCAASS a une meilleure résolution temporelle et spatiale que celle des MFCC car elle incorpore l'information de phase. De plus, les coefficients obtenus lors de la transformation d'analyse sont aisément inversibles car ils permettent de faire la transformation de synthèse facilement afin d'obtenir une reconstruction presque parfaite. Tous ces avantages laissent présager qu'il serait hautement profitable pour un système comme celui de Hoffman *et al.* [16] d'utiliser les LCAASS au lieu des MFCC pour optimiser la qualité de la reconstuction. Afin de valider cette hypothèse, nous avons mis en application la méthode LCAASS de transformation du signal avec la méthode stochastique nommée "Processus de Dirichlet hiérarchique et modèle caché de Markov" (HDP-HMM) de modélisation de la musique proposée par Hoffman *et al.* [16]. Cela fut réalisé lors de l'étude préliminaire ayant été conduite pour définir le projet de recherche. Les résultats démontrèrent une qualité de reconstruction nettement supérieure, telle qu'anticipée. Par contre, nous avons pris conscience des diverses limitations du LCAASS telles que la taille de la mémoire et la durée nécessaire au codage de longs signaux. En effet, pour coder de longs signaux, il est nécessaire d'augmenter le nombre et la taille des filtres servant à transformer le signal afin de conserver l'information de phase du signal dans le temps. Un signal long nécessite exponentiellement plus de mémoire en plus de nuire énormément à la vitesse de convergence du LCAASS.

Nous avons donc décidé de porter notre attention sur la réalisation du LCAASS plutôt que sur le système complet de composition musicale. En effet, sans une mise en oeuvre

efficace du LCAASS, il est difficile de songer à une utilisation intéressante d'un système de composition musicale qui utiliserait cette technologie.

Ainsi, ce mémoire porte sur la bonification des performances du "Locally competitive agorithm" (LCA) adapté aux signaux sonores (LCAASS), jusqu'à un codage et une reconstruction de qualité en temps réel, afin qu'il soit utilisable par diverses méthodes de composition musicale assistées par ordinateur (MCMAO). Pour ce faire, nous proposons une solution nommée "Fast Locally Competitive Algorithm for Audio" (FLCAA) capable de diminuer la durée du codage LCAASS et nécessitant moins de mémoire pour coder des signaux de longue durée. La différence majeure réside dans l'utilisation d'une technique découpant en morceaux de courte durée le signal à coder afin que chaque morceau soit codé individuellement. L'hypothèse motivant le FLCAA est qu'en n'ayant pas besoin de coder la phase dans le dictionnaire surcomplet comme pour le LCAASS, il serait possible de réduire grandement la durée du codage et l'utilisation de mémoire tout en conservant une bonne qualité de reconstruction. Si l'hypothèse est validée, il serait possible de coder de longs signaux rapidement en utilisant moins de mémoire. Ainsi, on pourrait remplacer le codage MFCC, et les LCAASS, afin de bonifier les performances de la MCMAO de Hoffman *et al.* [16]. Cela permettrait l'obtention de sorties de meilleure qualité et surtout plus rapidement.

1.2 Définition du projet de recherche

Le présent projet consiste à définir et à présenter le FLCAA, un système de codage parcimonieux permettant l'analyse perceptuelle des signaux sonores en temps réel. Il combine l'analyse par fenêtre coulissante avec l'algorithme LCA, de Rozell *et al.* [32]. L'implémentation du FLCAA sera évaluée en comparant sa représentation perceptuelle avec celle obtenue lors d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante. Également, les performances du FLCAA seront comparées au LCAASS afin d'avoir un aperçu de son potentiel. Cela sera effectué en générant, avec les 2 techniques, des représentations perceptuelles et en comparant la qualité de la reconstruction et la durée du codage sur des signaux de différentes durées. De plus, la robustesse du codage FLCAA est mise en évidence.

1.3 Objectifs du projet de recherche

L'objectif de ce mémoire est de présenter le FLCAA, une méthode d'analyse/synthèse qui permettrait de bonifier le MCMAO de Hoffman *et al.* [16] en remplaçant les MFCC. Suite aux recherches préliminaires, où l'on substitue les coefficients MFCC par les coefficients LCAASS, on identifie diverses limitations liées principalement au codage de la phase du signal dans le dictionnaire surcomplet. Ainsi, on désire éliminer complètement le codage de la phase dans le dictionnaire surcomplet. Pour ce faire, on propose d'incorporer un mécanisme analysant chaque segment indépendemment avec un certain chevauchement entre les segments. Cette technique est communément nommé analyse par fenêtre coulissante et permet de reconstruire le signal en conservant l'information de phase. De plus, comme le dictionnaire surcomplet est plus petit, cela permet de réduire la quantité de mémoire utilisée en plus de réduire la durée nécessaire au codage de longs signaux. En d'autres mots l'objectif du projet consiste à formaliser et à réaliser un système d'analyse/synthèse en temps réel, qui est exacte et robuste. Voici, ci-dessous, la liste des sous-objectifs permettant l'atteindre des objectifs principaux :

1. Formalisation du FLCAA afin d'inclure la technique d'analyse par fenêtre coulissante pour la transformation d'analyse, le codage parcimonieux et la transformation de synthèse.

2. Choix, génération et adaptation du banc de filtres afin d'obtenir la meilleure analyse et synthèse possible. Les coefficients obtenus suite à la transformation d'analyse doivent permettre la reconstruction du signal pour être utilisable par une MCMAO.

3. Implémentation du FLCAA telle que formalisée.

4. Validation de la fonctionnalité de l'implémentation FLCAA.

5. Évaluation de l'intérêt d'utiliser le FLCAA comme méthode d'analyse/synthèse du signal pour une méthode de composition musicale assistée par ordinateur.

6. Évaluation de la robustesse du FLCAA en omettant un pourcentage des coefficients obtenues lors du codage du signal pour réalisé la reconstruction. Cette omission correspond à une perte d'information et permet d'évaluer le bon fonctionnement de la reconstruction du système en cas d'erreur de transmission. En évaluant l'impact de cette perte d'information il sera possible de déterminer la limite de la robustesse afin de déterminer l'information minimale nécessaire pour obtenir un signal reconstruit qui est objectivement de bonne qualité.

7. Évaluation de la qualité des reconstruction FLCAA en comparant avec les reconstructions LCAASS selon le rapport signal sur bruit (RSB), le RSB segmentaire, la distorsion spectrale logarithmique (DSL) et l'évaluation perceptuelle de la qualité de la parole (EPQP).

8. Comparaison des durées nécessaires au codage du LCAASS et FLCAA, selon la durée du signal à coder, en tenant compte de la taille et du nombre de filtres cochléaires.

1.4 Contribution originale

La contribution originale principale de ce mémoire est le système FLCAA formalisé et fonctionnel. Le FLCAA est un système novateur générant un codage parcimonieux en temps réel permettant l'obtention d'une représentation perceptuelle qui offre une bonne résolution spatiale et temporelle des signaux sonores. Avec un tel système, il serait possible de bonifier diverses méthodes de composition assistées par ordinateur en fournissant plus d'informations sur le signal pour établir un modèle plus précis. Le FLCAA pourrait également être appliquée à diverses tâches de détection et/ou de reconnaissance sonore en temps réel. Une autre contribution est la comparaison de la qualité de reconstruction du FLCAA face aux LCAASS [29] afin de déterminer leurs performances relatives. Un FLCAA permettant une bonne qualité de reconstruction permettrait possiblement de bonifier le système de composition de Hoffman *et al.* [16] et autres MCMAO.

1.5 Plan du document

Suite à cette mise en contexte du mémoire, le chapitre intitulé État de l'art présente diverses MCMAO et converge vers l'identification de la problématique de ce mémoire. Ensuite, on présente diverses techniques de codage qui permettent de mettre en contexte le FLCAA venant répondre à la problématique. Cela, en débutant par une technique de filtrage par banc de filtres et en couvrant plusieurs autres technologies d'analyse/synthèse permettant l'obtention d'une représentation perceptuelle ainsi qu'une reconstruction de qualité des signaux sonores. La plupart des ouvrages présentés ont inspirés les présents travaux. Les plus déterminants furent le LCA [32] et son application pour les signaux sonores [29].

Le chapitre suivant intitulé FLCAA : Filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante et le codage LCA, présente une technique de reconstruction robuste du signal original sur des fenêtres d'analyse avec un haut taux de chevauchement. On y re-

trouve les explications, les schémas et les équations mathématiques définissant le FLCAA. De plus, les différences entre l'implémentation du LCA, du LCAASS et du FLCAA sont clairement expliquées. Finalement, la méthode de reconstruction est validée mathématiquement.

Par la suite, la section Expériences et conditions expérimentales présente la méthodologie et les paramètres utilisés pour évaluer et comparer correctement les sous objectifs posés précédemment. Dans la section Résultats et discussions on analyse l'implémentation et le potentiel du FLCAA. De plus, on compare les qualités de reconstruction du FLCAA et du LCAASS selon les critères de performance retenus. Le mémoire se termine en faisant la synthèse des informations pertinentes qui permettent de déterminer la qualité de l'analyse/synthèse FLCAA et sa capacité à bonifier une ou plusieurs méthodes de composition musicale assistées par ordinateur.

CHAPITRE 2

État de l'art

Ce chapitre permet d'établir un cadre de référence au FLCAA en plus d'exposer les technologies qui s'y rattachent.

Depuis la transformation de Fourier, permettant l'analyse d'un signal dans le domaine spectral, il a été démontré à maintes reprises qu'il est possible d'obtenir des représentations mettant en évidence certaines caractéristiques du signal en transformant un signal vers un autre domaine d'analyse. Par exemple, l'utilisation du spectrogramme permet d'obtenir une représentation fréquencielle ayant une très bonne résolution en fréquence. Cela a permis de pousser l'analyse des signaux en utilisant de l'information autrement difficile à obtenir. Depuis, plusieurs chercheurs ont mis en oeuvre diverses nouvelles techniques de transformation du signal à des fins d'analyse et/ou de codage. Toutefois, il existe beaucoup moins de techniques d'analyse qui permettent également une synthèse de bonne qualité du signal suite à l'analyse. Comme la synthèse du signal est nécessaire pour plusieurs MCMAO, on présente ci-dessous diverses transformations d'analyse/synthèse pertinentes menant à la compréhension du FLCAA.

2.1 Filtrage par banc de filtres

Le pré-traitement d'un signal a divers avantages lors de la modélisation de celui-ci. Lorsque bien effectué, il permet d'obtenir une représentation du signal exhibant plus de caractéristiques qui peuvent être utilisées pour établir un modèle plus performant. Afin d'utiliser un pré-traitement quelconque avant l'entraînement d'un modèle musical, il est nécessaire qu'il soit inversible. Cela, afin de permettre la reconstruction du signal suite à l'obtention des sorties du modèle musical. De façon générale, le filtrage par banc de filtres n'est pas inversible car il ne permet pas la reconstruction du signal suite à son codage. Les travaux de Vetterli [39], en 1986, présentent un banc de filtres inversible particulièrement intéressants car, il permet l'analyse, la synthèse et la compression des données par décimation et interpolation. Le filtrage du FLCAA est similaire à celui de Vetterli mais contrairement à lui, la compression du FLCAA s'effectue par obtention de parcimonie et non par décimation.

En plus de permettre une reconstruction parfaite, le codage par banc de filtres à M bandes ("M-band filterbank coding") de Vetterli analyse plus finement un signal sonore en utilisant plusieurs filtres au lieu d'un seul. Concrètement, tel que montré à la figure 2.1, il s'agit du filtrage d'un signal sonore par un ensemble de filtres passe-bande ($H_i(z)$) qui, après filtrage, permettent l'obtention de bandes spectrales distinctes. Chacune des bandes a une plage spécifique de fréquences définie par chacun des filtres passe-bande. Cela permet, par exemple, d'éditer le signal en modifiant seulement quelques bandes du signal ou de sélectionner les bandes principales afin d'effectuer diverses techniques de compression et de codage.

De façon plus spécifique, le filtrage par banc de filtres de Vetterli consiste à effectuer une transformation d'analyse $H_i(z)$ sur N filtres, suivie d'une décimation, permettant la compression du signal original $x(n)$, par un facteur $R \leq N$. Pour reconstruire le signal original $x(n)$, il suffit d'effectuer une interpolation d'un facteur R et N transformations de synthèse $F_i(z)$. La sommation des $F_i(z)$ obtenus permet de compléter la reconstruction du signal $\hat{x}(n)$. Pour le FLCAA, la transformation d'analyse $H_i(z)$ correspond à l'équation 3.7 et la transformations de synthèse $F_i(z)$ correspond à l'équation 3.10.

2.2 Banc de filtres cochléaires et représentation perceptuelle

Afin d'évaluer la qualité de la représentation perceptuelle du FLCAA, il est important de comprendre ce qu'est une représentation perceptuelle. À titre d'exemple, on pense aux travaux de Yang *et al.* [41] présententant, en 1992, un système qui filtre les signaux sonores à l'aide d'un banc de filtres cochléaires bio-inspirés du système auditif. Ils démontrent qu'il est possible d'obtenir une représentation cochléaire suite au filtrage d'un signal sonore par un banc de filtres cochléaires.

Également, ils concluent que "*L'examen détaillé des représentations auditives révèle une amplification des caractéristiques du signal ainsi qu'une meilleure résistance au bruit* [41]". Ainsi, on constate que les représentations auditives sont mieux adapté pour l'analyse des signaux sonores que les représentations traditionnelles. Cela renforce la validité d'un choix de bases bio-inspirées pour le FLCAA comme pré-traitement pour mettre en évidence les caractéristiques du signal sonore à modéliser. Yang *et al.* [41] proposent également un mécanisme d'inhibition latérale non itératif contrairement à celui du FLCAA.

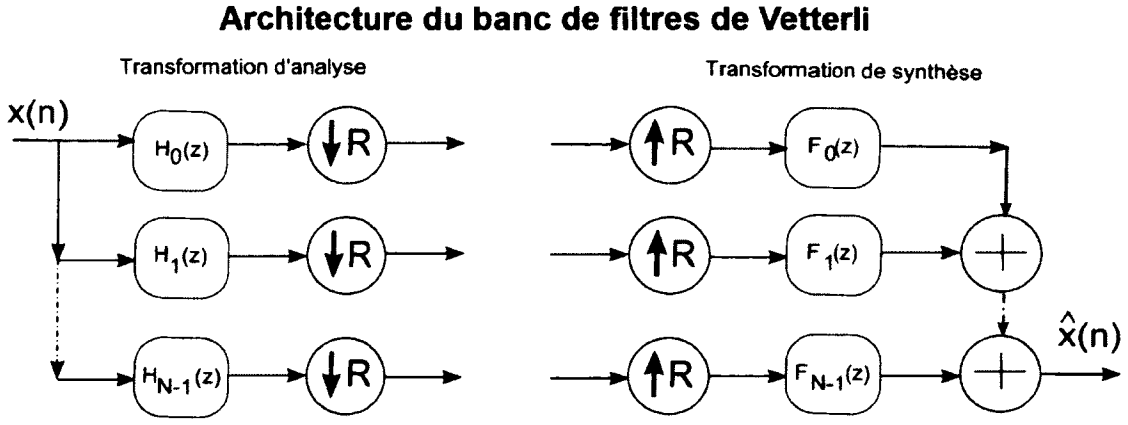

Figure 2.1 Architecture du banc de filtres de Vetterli [39] ayant un nombre N de filtres. R correspond au facteur de compression, de décimation et d'interpolation qui doit être inférieur ou égal à N pour favoriser une bonne qualité de reconstruction. La décimation consiste à éliminer des échantillons sonores du à intervalle régulier afin de réduire la fréquence d'échantillonnage du signal. Par exemple, un signal échantillonné à 32KHz devient échantillonné à 8KHz si $R = 4$ et est compressé du trois quart. La décimation entraîne une perte des fréquences élevées du signal selon le théorème de Nyquist-Shannon. Inversement, l'interpolation consiste à estimer des échantillons sonores et les insérer dans le signal décimé afin d'augmenter la fréquence d'échantillonnage du signal. Contrairement aux travaux de Vetterli utilisant la décimation comme stratégie de compression suite au filtrage $H_i(z)$, le FLCAA utilise la parcimonie comme stratégie de compression suite au filtrage décrit par l'équation 3.7. La parcimonie correspond à la sélection des filtres les plus importantes pour la représentation du signal (équation 2.5). Les autres filtres sont posés à zéro ce qui facilite la compression et le codage du signal. Pour effectuer la reconstruction du signal Vetterli utilise une interpolation égale à R suivi d'une transformation de synthèse $F_i(z)$. Le FLCAA ne nécessite pas de traitement spécifique avant d'effectuer la transformation de synthèse définie par l'équation 3.10 car la parcimonie, contrairement à la décimation, n'a pas à être inversé pour obtenir la reconstruction.

2.3 Algorithme localement compétitif ou "Locally Competitive Algorithm" (LCA)

Dans cette section nous décrivons le "Locally Competitive Algorithm" (LCA) présenté par Rozell *et al.* [32] car ces travaux ont donné lieu au LCAASS et subséquemment à notre FLCAA. En effet, le pré-traitement et les stratégies itératives d'optimisation du LCAASS et du FLCAA sont les mêmes que celles de Rozell *et al.*. Ce sont le signal d'entrée et le banc de filtres cochléaires qui diffèrent principalement entre le LCA, le FLCAA et le

LCAASS. Les différences de l'implémentation LCAASS qui adapte le LCA afin qu'il puisse coder un signal sonore sont exposées en détails à la section 2.4. De même, les différences de l'implémentation FLCAA qui adapte le LCAASS en introduisant un mécanisme de fenêtre coulissante sont exposées en détails au chapitre 3.

Le LCA est une méthode de codage parcimonieux constitué d'un seuillage et de compétition entre les neurones. Le codage est parcimonieux lorsqu'il exhibe peu de coefficients non nul. Cela permet la selectivité des caractéristiques importantes du signal. LCA montre un excellent potentiel pour le traitement d'images, sa convergence est assurée mathématiquement [3] et est aisément implémentable sur ordinateur. De plus, selon Rozell *et al.* [32], "*le LCA s'exécute plus rapidement que le "Matching Pursuit(MP)" [20] et les autres méthodes d'optimisation gloutones ("Greedy") [38] effectuant un codage parcimonieux*". Cela, car le LCA oriente la recherche d'une solution optimale à l'aide de la charge interne des neurones utilisés ce qui permet d'atteindre une solution localement optimale avec beaucoup moins de calculs, donc plus rapidement, que pour le MP ou autres méthodes d'optimisation gloutones. Également, le mécanisme de seuillage accélère la convergence de l'algorithme vers une solution optimale.

LCA débute par la transformation d'un ensemble de pixels, extraits de l'image d'entrée, en le projetant sur un ensemble de bases bio-inspirées du système visuel un peu comme Yang *et al.* [41] l'ont fait en 1992 dans le contexte de signaux sonores. Toutefois, contrairement aux travaux de Yang *et al.* utilisant les ondelettes, le LCA recourt aux filtres de Gabor conçus pour le codage d'images afin de transformer le signal. Les filtres de Gabor sont des filtres linéaires permettant de détecter les contours. Ils permettent, entre autre, l'extraction des caractéristiques d'une image. Rozell, voulant établir un modèle physiologique du système visuel, les a choisi pour le LCA étant donné que les cellules du cortex visuel des mammifères sont modélisables par des filtres de Gabor [10] [11]. Au niveau physiologique, les champs récepteurs des groupes de neurones correspondent aux filtres de Gabor. Les réponses de ceux-ci au signal analysé correspondent au taux moyen de décharge des groupes de neurones répondant plus ou moins fortement au signal reçu.

Le résultat du filtrage est ensuite fourni au système itératif d'optimisation de la parcimonie de la solution. Pendant les itérations, LCA tente de minimiser l'erreur quadratique moyenne (EQM) entre le signal et sa reconstruction. Pour ce faire, il utilise simultanément deux mécanismes bio-inspirés permettant l'obtention d'une solution parcimonieuse : l'inhibition latérale et le seuillage.

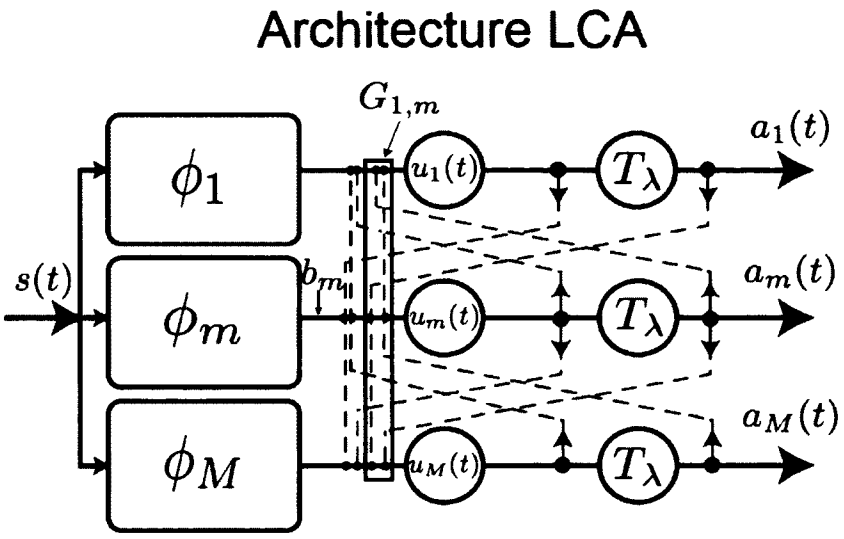

Figure 2.2 Architecture du système LCA tel que proposé par Rozell *et al.* [32]. Chaque champ récepteur ϕ_m est une implémentation d'un filtre de Gabor conçu pour le codage d'images. $a_m(t)$ est la sortie du seuillage T_λ des potentiels internes $u_m(t)$ pour chaque neurone m (équation 2.3). Suite au filtrage initial par banc de filtres cochléaires (équation 2.4) permettant d'obtenir b_m, les neurones m ayant des ϕ_m similaires s'inhibent entre eux par inhibition latérale selon G (equation 2.2). L'inhibition latérale, encadrée par un rectangle noir, s'applique aux coefficients $a_m(t)$ à chaque itération afin de mettre à jour les potentiels internes $u_m(t)$ (équation 2.5). $u_m(t)$ est également utilisé à chaque itération pour sa propre mise à jour.

Le mécanisme d'inhibition latérale s'effectue à chaque itération en projetant les coefficients LCA $a_j(t)$, de chaque neurone j, sur la matrice d'inhibition latérale G tel que :

$$\sum_{j \neq i}^{M} G_{i,j} a_j(t) \qquad \text{pour i et j} = 1, ..., \text{M} \tag{2.1}$$

Cela favorise l'atteinte rapide d'une solution optimale en orientant les recherches de LCA. La matrice d'inhibition latérale, de taille $M \times M$, est définie par l'équation suivante.

$$G_{i,j} = \phi_i^t \phi_j \qquad \text{pour i et j} = 1, ..., \text{M} \tag{2.2}$$

$G_{i,j}$ est donc un scalaire qui représente la corrélation entre les 2 champs récepteurs. G encourage la parcimonie de la solution en inhibant les filtres similaires entre eux pour minimiser la redondance de l'information à la sortie de ceux-ci.

Le mécanisme de seuillage s'effectue à chaque itération sur les potentiels internes $u_m(t)$, de chaque neurone m. Cela permet de générer les coefficients $a_m(t)$ tel que défini par la formule suivante.

$$a_m(t) = T_\lambda(u_m(t)) \tag{2.3}$$

Comme le potentiel interne correspond à l'activité neuronale, le seuillage agit en éliminant à chaque itération les réponses neuronales, ou coefficients $u_m(t)$, sous le seuil λ. Cela, combiner à l'inhibition latérale, favorise la convergence et la parcimonie du codage en diminuant le nombre de neurones dont les sorties, représentées par les $a_m(t)$, sont différentes de zéro. LCA itère jusqu'au nombre maximal d'itérations posé ou jusqu'à l'atteinte d'une solution ayant une reconstruction de qualité suffisante pour atteindre le niveau de distortion maximale posé. La convergence de LCA est assuré [3] et l'utilisation d'un seuillage dur (Figure 2.3) correspondant à la norme mathématique L0 qui assure l'atteinte d'une solution globalement optimale tandis que, l'utilisation d'un seuillage mou correspond à la norme mathématique L1 et assure l'atteinte d'une solution localement optimale.

Au niveau physiologique, tel que modélisé par le codage LCA, on observe une augmentation de la parcimonie des réponses neuronales lors de l'intégration visuelle due à l'inhibition latérale et au seuillage [40].

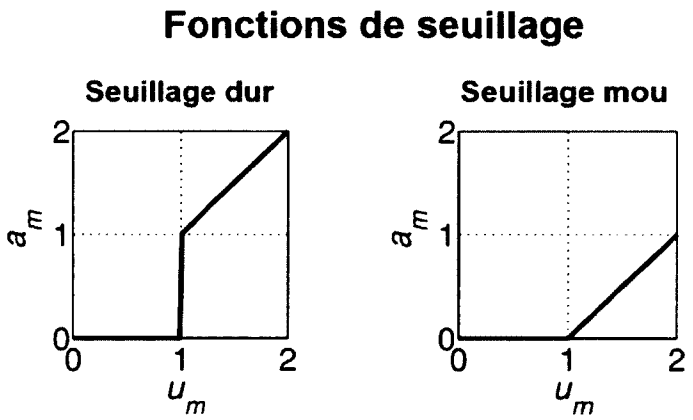

Figure 2.3 Fonctions de seuillage, présenté par Rozell [32], établissant la relation entre le potentiel interne $u_m(t)$ et la sortie du seuillage $a_m(t)$ correspondant pour chaque neurone m. L'image de gauche correspond à un seuillage dur et assure la convergence de l'algorithme selon la norme mathématique L0 ce qui permet d'obtenir une solution globalement optimale. L'image de droite correspond à un seuillage mou et assure la convergence de l'algorithme selon la norme mathématique L1 ce qui permet d'obtenir une solution localement optimale.

2.3.1 Description détaillée de l'algorithme LCA

Transformation d'analyse

Un ensemble Φ de M bases (ou filtres) ϕ_m est premièrement défini. Il est possible d'utiliser plusieurs ensembles de bases différents. Le choix des bases est laissé à la discrétion du concepteur de l'algorithme. Chaque base ϕ_m est représentée par un neurone dont l'indice est m. Le champ récepteur du neurone m correspond à ϕ_m. Les modèles de neurones sont de type intégration du potentiel interne du neurone produisant une variable de sortie non nulle si le potentiel atteint un seuil minimal. C'est donc un modèle mixte intermédiaire entre le neurone à intégration et décharges et un modèle encodant le rythme de décharges. Le codage résultant est l'ensemble des coefficients a_m de la représentation parcimonieuse. Celui-ci représente le taux moyen de décharges pour chaque neurone m(figure 2.2). L'équation suivante décrit la transformation d'analyse obtenue en faisant la projection du signal d'entrée $s(t)$ sur la transposée du champ récepteur ϕ_m^t de chaque neurone m.

$$b_m = \phi_m^t s(t) \tag{2.4}$$

Avec $s(t) = [s_1(t), s_2(t)\ , ..., s_L(t)]^t$, le vecteur colonne représente le signal au temps t, et $\phi_m = [\phi_{1,m}, \phi_{2,m}, .., \phi_{N,m}]^t$, le vecteur colonne représente le champ récepteur du neurone m.

Optimisation intérative

Chaque itération du LCA optimise le résultat du codage parcimonieux $a_m(t)$ (ou coefficients LCA). Pour ce faire, LCA combine la fonction de seuillage T_λ et la matrice d'inhibition latérale G. L'amplitude de $b_m(t)$ représente le degré de similarité entre le signal $s(t)$ et le champ récepteur du neurone m. On obtient $a_m(t)$ suite au seuillage T_λ tel qu'expliqué à la section précédente. L'évolution des coefficients LCA $a_m(t)$ et des potentiels internes $u_m(t)$ pour tous les neurones m dans le temps est régie par l'équation ci-dessous.

$$\frac{du_m}{dt} = \frac{1}{\tau}\left[b_m(t) - u_m(t) - \sum_{j \neq m}^{M} G_{m,j} a_j(t)\right] \tag{2.5}$$

Le facteur d'inertie τ est habituellement posé à 0.01. La sommation $\sum_{j \neq m} G_{m,j} a_j(t)$ représente le résultat des contributions pondérées de tous les neurones connectés au neurone m alors que $G_{m,j}$ (equation 2.2) correspond au niveau de similarité entre les champs récepteurs des neurones m et j.

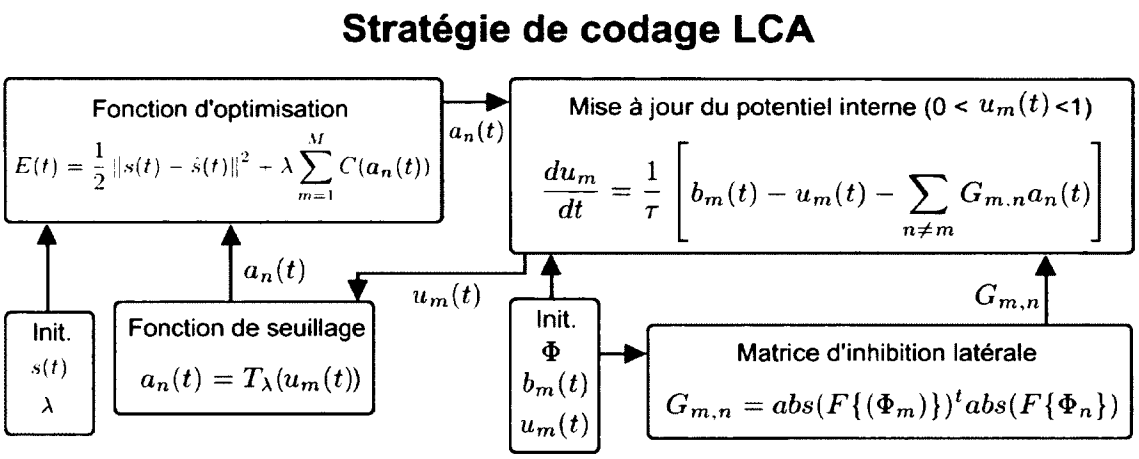

Figure 2.4 Stratégie de codage LCA. Une représentation parcimonieuse $(a_n(t), n = 1, ..., M)$ du signal d'entrée est évaluée itérativement. Chaque itération de la fontion d'optimisation met à jour le potentiel interne $u_m(t)$ de chaque neurone pour $m = 1, ..., M$. L'algorithme améliore la parcimonie de la solution au cours des itérations grâce à la combinaison de l'effet de la fonction de seuillage T_λ et de la matrice d'inhibition latérale G qui est déterminé à l'aide du produit des transformée de Fourier rapide F du banc de filtres Φ et de sa transposée. G correspond au niveau de similarité des filtres entre eux.

Critères d'optimisation du codage

Afin d'optimiser la parcimonie du codage, LCA tend vers zéro la fonction d'énergie $E(t)$ en combinant l'erreur quadratique moyenne de reconstruction à une fonction de coût C afin d'obtenir un codage parcimonieux permettant une bonne qualité de reconstruction. En faisant tendre $E(t)$ vers zéro, on optimise la parcimonie de la solution car, plus il y a de coefficients non nuls, plus $E(t)$ est élevé. La fonction de coût C permet la parcimonie du codage et est déterminé en fonction de la fonction de seuillage T_λ sur le potentiel interne u_m pour chaque neurone m. Cette relation est présenté par l'équation 2.6.

$$\lambda \frac{dC(a_m)}{da_m} = u_m - a_m = u_m - T_\lambda(u_m) \tag{2.6}$$

En fait, LCA code un signal $s(t)$ quelconque par une approximation itérative des coefficients parcimonieux $a_m(t)$ (pour $m = 1, ..., M$) avec les contraintes que le signal reconstruit $\hat{s}(t)$ doit être le plus proche possible du signal original $s(t)$ et que les coefficients $a_m(t)$

soient le plus parcimonieux possible.

$$E(t) = \frac{1}{2} \|s(t) - \hat{s}(t)\|^2 + \lambda \sum_{m=1}^{M} C(a_m(t)) \tag{2.7}$$

Ainsi, on tente de minimiser la fonction d'énergie $E(t)$ à l'aide de la fonction de coût C. On valide la qualité du codage et de la reconstruction avec l'erreur quadratique moyenne de reconstruction $\frac{1}{2} \|s(t) - \hat{s}(t)\|^2$. La minimisation de la fonction d'énergie est assurée par le mécanisme de seuillage qui élimine les coefficients faibles.

Transformation de synthèse

Il est possible d'obtenir le signal reconstruit $\hat{s}(t)$ en projetant les coefficients résultant du codage LCA $\mathbf{a}(t) = [a_1, a_2, ..., a_M]$ sur le dictionnaire surcomplet $\Phi = [phi_1, phi_2, ..., phi_M]$. Plus précisément, $\hat{s}(t)$ est obtenu en pondérant l'ensemble des bases ϕ_m avec les coefficients a_m pour $m = 1, ..., M$.

$$\hat{s}(t) = \sum_{m=1}^{M} a_m(t)\phi_m = \Phi \mathbf{a}(t) \tag{2.8}$$

2.4 Application du LCA pour le codage de signaux sonores

En 2010, Pichevar *et al.* [30] présentent une nouvelle technique de représentation sonore perceptuelle. Il s'agit d'une adaptation de l'algorithme LCA de Rozell *et al.* [32] pour les signaux sonores. Celle-ci présente un bon niveau de parcimonie et une bonne qualité de reconstruction. La parcimonie entre les filtres a comme effet d'amplifier les caractéristiques du signal en concentrant l'information sur les coefficients d'un nombre restreint de filtres. Concrètement, en éliminant les coefficients faibles et en concentrant l'information sur un nombre minimale de filtres, il est plus facile d'observer les fréquences importantes du signal dans le temps parce qu'il y a moins d'informations non pertinentes qui nuisent à l'analyse de la représentation. Cela est observable sur la figure 2.5 qui correspond à la représentation perceptuelle du LCAASS de Pichevar *et al.*. On y observe une solution parcimonieuse, avec peu de coefficients différents de zéro, qui met en évidence les caractéristiques du signal. Une force importante de cette technique, afin d'être intégré à une MCMAO, est la simplicité de la synthèse par une simple projection du codage sur la transposée des réponses impulsionnelles des filtres cochléaires.

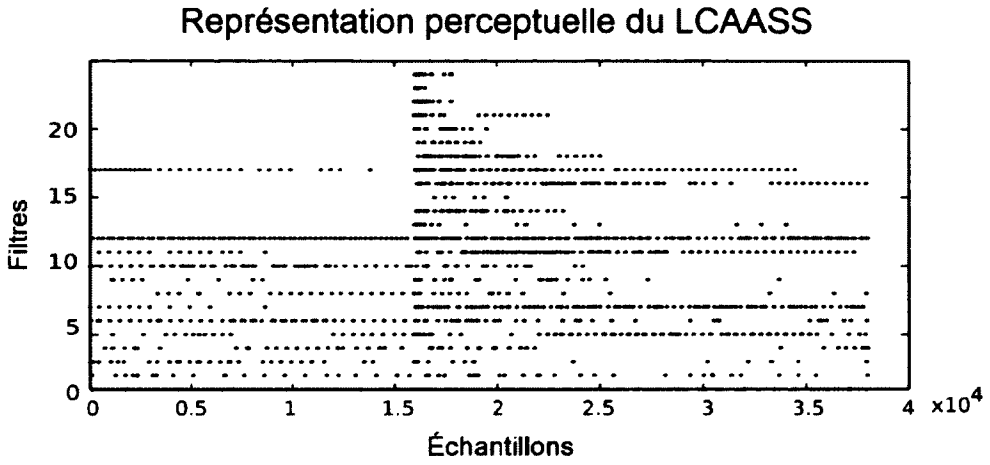

Figure 2.5 Représentation auditive perceptuelle du système LCAASS présenté par Pichevar *et al.* [30]. L'ordonnée est associée aux 24 filtres cochléaires Gammatone utilisés. Ainsi, le champ récepteur de chaque neurone est associé à un filtre cochléaire Gammatone. L'abscisse est exprimé en échantillons et correspond à l'échelle du temps discrétisé. On observe que la représentation perceptuelle est parcimonieuse et que les fréquences importantes sont mise en évidence en fonction du temps.

Plus précisément, LCAASS utilise les réponses impulsionnelles de $K = 24$ filtres de type gammatone en tant que bases ϕ_m. Cela correspond à encoder le signal sur 24 canaux cochléaires. De plus, chaque base ϕ_m est décalée de q échantillons pour créer une autre base identique, mais décalée dans le temps. Le dictionnaire surcomplet Φ, utilisé par le système LCAASS, est défini en regroupant les K bases avec leurs versions décalées dans le temps. Cela permet d'encoder directement l'information de phase dans Φ et de conserver la phase afin de permettre une reconstruction de qualité. Ainsi, le dictionnaire Φ comprend $K \times L/q$ bases, avec L étant la longueur en échantillons du signal original $s(t)$, tel que montré par l'équation 2.9.

$$\Phi_n^t = \begin{bmatrix} \Phi_{n,1}\Phi_{n,2}\Phi_{n,3}\cdots\Phi_{n,m}\cdots 0 \\ 0_{1\times q}\Phi_{n,1}\Phi_{n,2}\cdots\Phi_{n,m-q}\cdots 0 \\ 0_{1\times q}0_{1\times q}\Phi_{n,1}\cdots\Phi_{n,m-2*q}\cdots 0 \\ \vdots\ddots\vdots\ddots\vdots \\ 0_{1\times q}0_{1\times q}0_{1\times q}\cdots\Phi_{n,1}\cdots\Phi_{n,m} \end{bmatrix} \tag{2.9}$$

Malgré son excellent potentiel, LCAASS souffre de deux limitations majeures. La première est une limitation de mémoire étant donné qu'on code la phase directement dans le

dictionnaire surcomplet Φ. Cela force un compromis entre la qualité de reconstruction et la longueur du signal à coder, car Φ croît rapidement lorsque la longueur du signal $s(t)$ croît. La seconde limitation est le temps d'exécution. Le LCAASS nécessite un bon nombre d'itérations pour converger afin d'assurer une bonne qualité de reconstruction. Cela rend cette technique inutilisable pour l'analyse/synthèse de signaux sonores en temps réel, spécialement pour les longs signaux.

Ainsi, la réduction de la durée des fenêtres d'analyse semble indiquée pour contrer la surutilisation de la mémoire et la lenteur de la convergence de l'algorithme. Toutefois, selon nos observations, même en codant de courtes fenêtres d'analyse et en réalisant une reconstruction par chevauchement et addition, "Overlap and add", il n'est pas possible d'obtenir un codage temps réel. En fait, si on diminue trop la taille de la fenêtre d'analyse il y à dégradation du signal reconstruit. Lorsque l'on utilise une fenêtre de taille minimale mais permettant une bonne reconstruction on constate que le temps nécessaire au codage est plus lent que temps réel.

2.5 Causal Local Competitive Algorithm (CLCA)

Récemment, Charles *et al.* [6] ont proposés le "Causal Local Competitive Algorithm" (CLCA), une représentation perceptuelle combinée à l'inhibition latérale dans le temps. Cela est fait sur des blocs de données par fenêtre d'analyse coulissante avec chevauchement. Ainsi, il semble possible d'inclure le concept de fenêtre coulissante avec le codage LCA. Bien que le CLCA soit une technique de compression prometteuse, la codage est lent, en raison du mécanisme d'inhibition latérale dans le temps, et est donc difficilement intégrable à une MCMAO. C'est pourquoi il n'y a pas eu d'implémentation du CLCA dans le cadre de ce mémoire. Il serait très intéressant de comparer ses performances au FLCAA dans des travaux futurs.

2.6 Réflexion sur l'état de l'art

En faisant l'étude de l'état de l'art, on arrive à la conclusion que ces techniques sont difficiles à intégrer à des MCMAO. Soit parce qu'ils sont trop lents, qu'ils ne peuvent pas analyser des signaux réels de longues durées, qu'ils sont avares en mémoire ou que la représentation perceptuelle est difficilement interprétable à cause de leur faible résolution. Aucun des travaux consultés ne fait mention d'une utilisation temps réel sur des signaux sonores dynamiques. Ainsi, on conclut qu'aucun des systèmes de représentation percep-

tuelle actuels ne possède une bonne applicabilité et qu'ils sont difficilement utilisables dans une MCMAO.

On définit l'applicabilité du système de représentation perceptuelle par la combinaison de la qualité de la reconstruction, la durée nécessaire au codage, la capacité à traiter de longs signaux, la capacité à définir un nombre variable de filtres cochléaires et la robustesse lors de la synthèse du signal. Le FLCAA se propose de pallier à ces limitations en intégrant tous ces aspects.

CHAPITRE 3

FLCAA : Filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante et le codage LCA

Le système FLCAA est un filtrage par banc de filtres cochléaires combinant une fenêtre coulissante et le codage LCA. Le LCA est expliqué en détail à la section 2.3. Le FLCAA se distingue par sa capacité de codage perceptuel des signaux sonores en temps réel. Il peut être décrit comme une technique d'analyse perceptuelle et de synthèse des signaux sonores, en temps réel, utilisant un banc de filtres cochléaires pour la transformation du signal. De plus, il utilise les mécanismes de seuillage et d'inhibition latérale pour optimiser la parcimonie du codage ainsi qu'une fenêtre d'analyse coulissante avec un haut taux de chevauchement fixé.

Tel que'expliqué à la section 2.4, la taille du dictionnaire surcomplet Φ est la principale limitation pour mettre en application le système LCAASS. Pour pallier à cette limitation, il semble approprié de réduire la taille de Φ au maximum en faisant correspondre à la durée discrétisée de la réponse impulsionnelle des filtres cochléaires avec le nombre de colonnes. Toutefois, comme la réponse impulsionnelle de notre implémentation Roexp [19] d'un filtre cochléaire est fenêtré, le résultat du filtrage est également fenêtré ce qui entraîne une perte d'information aux extrémités de la fenêtre. Cela implique, qu'il n'est pas possible d'accélérer LCAASS en définissant Φ tel un banc de filtres cochléaires pour ensuite reconstruire le signal à l'aide du chevauchement additif (Overlap and add) car, il y a perte d'information causée par troncature des extrémités des réponses impulsionnelles dans le temps. En effet, un chevauchement de 50%, typique des techniques d'analyse/reconstruction par chevauchement additif est inadéquat pour une l'obtention d'une bonne qualité de reconstruction. Néanmoins, nous avons découvert qu'en utilisant un chevauchement d'environ 90% entre les fenêtres d'analyse successives, on conserve l'information de phase, ce qui permet d'obtenir une bonne reconstruction ainsi qu'une représentation perceptuelle du signal analysé. Ainsi, on accélère grandement la convergence du LCAASS en codant l'information de phase à l'aide du chevauchement des fenêtres d'analyse au lieu que la phase soit codée directement dans Φ.

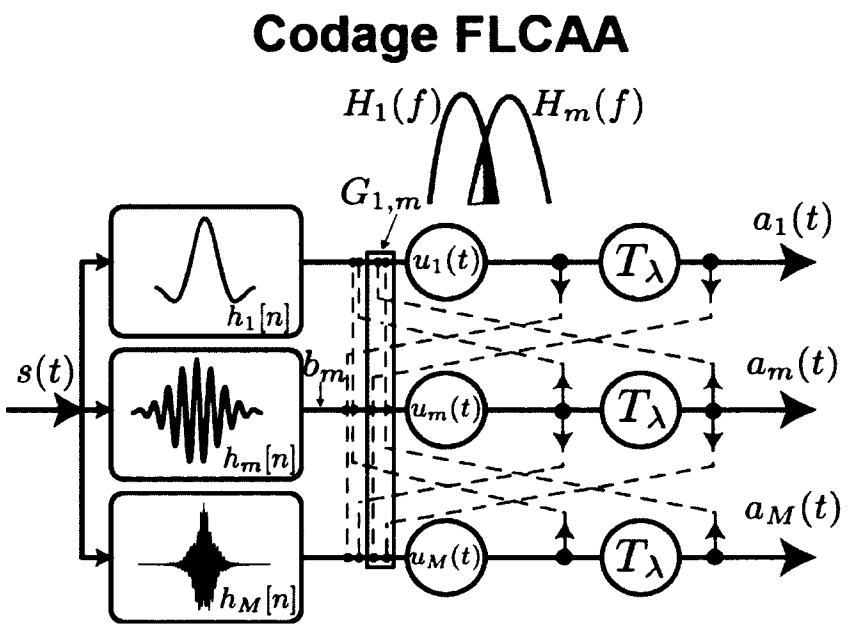

Figure 3.1 Architecture du réseau de neurones à inhibition latérale utilisé par le FLCAA. Chaque champ récepteur ϕ_m est la réponse impulsionnelle finie $h_m[n]$ d'un filtre "rounded exponential" (roExp) [19]. Les neurones s'inhibent entre eux selon G correspondant à la similarité des filtres dans le domaine spectral (équation 3.8). $a_m(t)$ est la sortie du codage obtenu après seuillage T_λ du potentiel interne $u_m(t)$ pour chaque neurone m. L'inhibition latérale G, encadrée par un rectangle noir, s'applique aux coefficients $a_m(t)$ à chaque itération afin de mettre à jour les potentiels internes $u_m(t)$ (équation 2.5). $u_m(t)$ est également utilisé à chaque itération pour sa propre mise à jour.

3.1 Modifications apportées au système LCA

3.1.1 Fenêtrage

On incorpore une fenêtre d'analyse coulissante avec un haut taux de chevauchement fixé au dessus de 90% pour la transformation du signal. Il y a une différence d'implémentation si l'on traite un signal temps réel ou un signal fini. Dans le cas où le signal est fini, on définit d'abord le nombre de fenêtres d'analyse L_w à l'aide de la durée L du signal $s(t)$ et de la durée N de la fenêtre d'analyse.

$$L_w = \lfloor (L - N)/l \rfloor . l \geq 1 \tag{3.1}$$

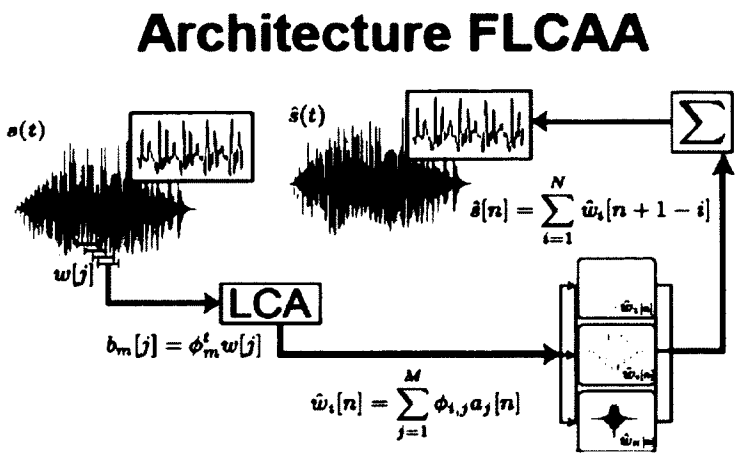

Figure 3.2 Architecture du système FLCAA, avec 128 filtres cochléaires, pour l'analyse et la synthèse de signaux sonores. Le signal $s(t)$ et une partie agrandie de celui-ci sont illustrés dans le coin supérieur gauche. Le signal est divisé en fenêtres w qui se chevauchent. Celles-ci sont projetées sur les filtres ϕ_m^t (Eq.3.7). Le résultat b_m est utilisé lors du codage en tant qu'objectif afin de générer un codage parcimonieux $a[n]$. Pour réaliser la reconstruction, chaque fenêtre d'analyse $\hat{w}$ doit être reconstruite préalablement (Eq.3.10). La reconstruction complète est ensuite obtenue grâce à l'équation 3.11. La reconstruction $\hat{s}(t)$ et une partie agrandie de la reconstruction sont illustrées en haut au milieu de la figure. $\hat{s}(t)$ a été obtenu en utilisant seulement une fenêtre $\hat{w}_i$ sur 4 lors de la reconstruction.

$\lfloor . \rfloor$ indique qu'on conserve uniquement la partie entière. l est le nombre d'échantillons de décalage entre deux fenêtres d'analyse successives. Ainsi, l doit être plus grand ou égal à un et s'il égal à un, le chevauchement est maximal. Sinon, le chevauchement décroit proportionnellement à l'augmentation de l.

Dans le cas où l'on traite un signal temps réel, on adapte L_w pour compter le nombre de fenêtres à analyser $w[j]$ reçues. Il est possible de mettre en mémoire tampon un groupe de $w[j]$ afin de les coder ensemble et ainsi pouvoir paralléliser le traitement. Il serait intéressant de déterminer la taille de la mémoire tampon optimale pour un traitement temps réel mais ce n'est pas traité dans le présent mémoire.

$$w[j] = s[(j-1)l+1], s[(j-1)l+2], ..., s[(j-1)l+N] \text{ pour j } = 1, ..., L_w \qquad (3.2)$$

On définit L_w fenêtres à analyser ($w[j]$) à l'aide de l'équation 3.7. Les $w[j]$ sont en fait des segments du signal $s(t)$ qui se chevauchent tel que présenté sur la gauche de la figure 3.2.

3.1.2 Conception des filtres cochléaires utilisés

Des filtres cochléaires, passe-bande, à pentes exponentielles arrondies, ou "rounded exponential" (RoExp), ont été choisis pour effectuer la transformation d'analyse. D'après les observations de Patterson [28] et Moore [25], les filtres auditifs sont similaires aux filtres RoExp [19]. Ils sont donc un bon choix pour favoriser l'extraction des caractéristiques d'un signal sonore. Ces filtres cochléaires sont caractérisés par la largeur de leurs bandes critiques, calculée à l'équation 3.3, et par la forme des exponentielles arrondies présentée à l'équation 3.4.

$$ERB(f_c) = 6,23f_c^2 + 93,39f_c + 28,52 \tag{3.3}$$

Où f_c est la fréquence centrale d'un filtre, exprimée en kHz, et ERB est la largeur de la bande critique, exprimé en Hz.

$$W(g) = (1 + pg)e^{-pg},\ g = |f - f_c| / f_c,\ p = 4f_c/ERB(f_c) \tag{3.4}$$

Où $W(g)$ est la fonction de trasfert du filtre, g est la déviation de la fréquence à partir du centre du filtre, divisée par la fréquence centrale f_c, et p détermine la largeur de la bande du filtre. Liu [19] présente la forme typique des filtres résultants sur la figure 3.3.

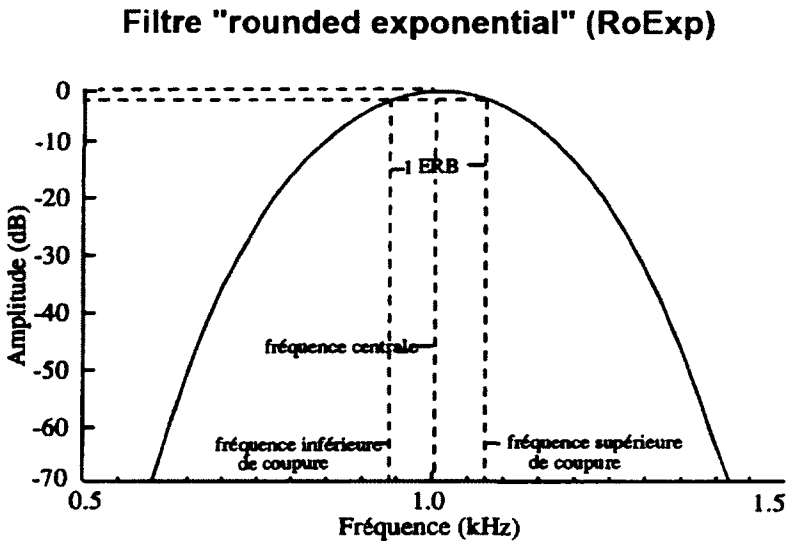

Figure 3.3 Forme schématisée d'un filtre auditif avec fréquence centrale F_c = 1008 Hz telle que présentée par Liu [19].

La figure 3.4 permet d'observer que la distribution des filtres est linéaire sur l'échelle logarithmique des fréquences.

Le choix du banc de filtres est justifiable par deux principes permettant le bon fonctionnement du FLCAA. Premièrement, l'implémentation RoExp [19] des filtres cochléaires produit des fonctions de transfert linéaire en phase ce qui permet aux coefficients parcimonieux $a_m(t)$ d'être utilisés directement en tant que représentation sonore perceptuelle. Deuxièmement, le banc de filtres est orthogonal symétrique ce qui permet d'effectuer la transformation de synthèse simplement. Les transformations d'analyse et de synthèse sont expliquées dans les sections suivantes.

Concrètement, chaque champ récepteur ϕ_m est la réponse impulsionnelle finie $h_m[n]$ d'un filtre RoExp. La génération des $h_m[n]$ a été réalisée dans les travaux Liu [19] et de Loiselle *et al.* [31] que nous utilisons afin de générer notre banc de filtres et est présenté à l'équation 3.5.

$$h[n] = \frac{1}{f_s} \int_{-\frac{f_s}{2}}^{\frac{f_s}{2}} H(f) e^{\frac{i2n\pi f}{f_s}} df \tag{3.5}$$

Où f_s est la fréquence d'échantillonnage et $h(n)$ la réponse impulsionnelle.

En fait, nous générons plus de filtres que nécessaire et conservons certains filtres permettant l'assemblage d'un banc de filtres ayant une distribution similaire à celle du système auditif (Figure 3.4), c'est-à-dire un fort chevauchement des filtres cochléaires en basse fréquence. De plus, afin d'assurer une contribution uniforme des filtres cochléaires, le banc de filtres est normalisé pour que la somme du gain des filtres soit égale à 1 pour toutes les fréquences. La normalisation est donnée par l'équation 3.6.

$$\sum_{i=1}^{M} |H_i(f)| = 1 \quad \forall f \tag{3.6}$$

3.1.3 Transformation d'analyse

FLCA utilise un décalage fixe de l échantillons entre les fenêtres d'analyse successives dans le temps. Chaque fenêtre d'analyse, ayant N échantillons, est projetée sur M bases de la matrice Φ (équation 2.4). Ensuite, chaque b_m résultant est codé par la méthode LCA [32] afin d'obtenir les coefficients a_m. La figure 3.2 présente la combinaison du LCA avec l'analyse par fenêtre coulissante. Voici l'équation 2.4 qui a été modifiée pour inclure

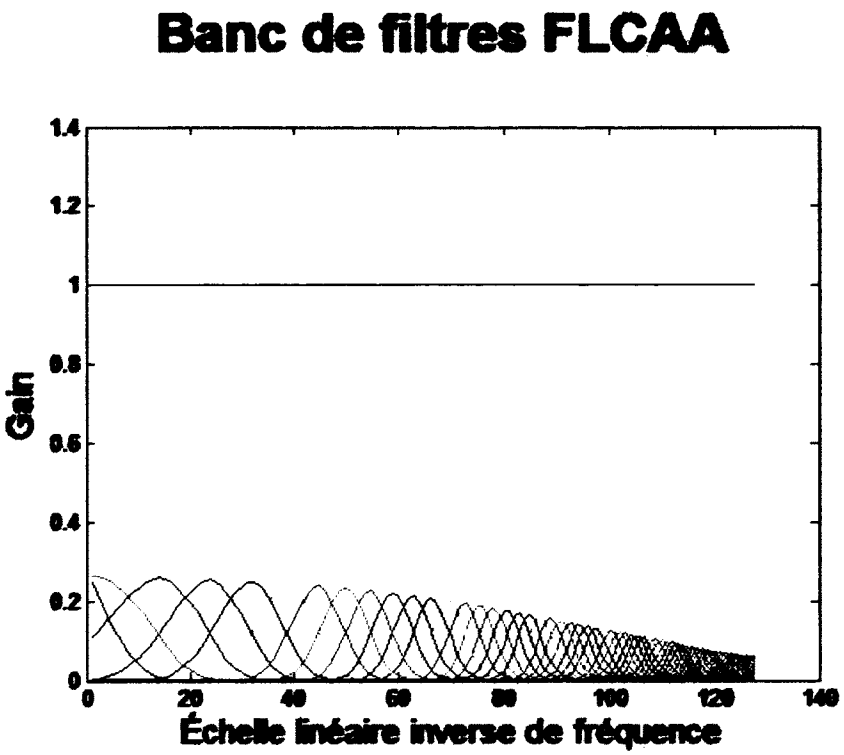

Figure 3.4 Représentation spectrale du banc de filtres utilisé par le FLCAA, le LCAASS et le filtrage par fenêtre coulissante. On présente, en abscisse, les fréquences centrales des 128 filtres cochléaires sur l'échelle linéaire inverse des fréquences, soit de 100 Hz à 7900 Hz. Ainsi, 1 sur l'échelle de fréquence correspond à 7900 Hz et 128 à 100 Hz. La distribution des filtres montre un fort chevauchement en basse fréquence, aussi observé dans le système auditif. La ligne bleue correspond à la sommation du gain des filtres, exprimée en ordonnée, pour chaque fréquence.

le chevauchement des fenêtres d'analyse.

$$b_m[j] = \phi_m^t w[j] \text{ pour j } = 1,..., L_w \tag{3.7}$$

$b_m[j]$ est le résultat de la projection de la fenêtre d'analyse j sur la base m et correspond à une fenêtre de représentation perceptuelle. Pour chaque fenêtre j, on définit chaque $b_m[j]$ permettant d'appliquer l'inhibition latérale et le seuillage qui augmentent la parcimonie du codage.

3.1.4 Inhibition latérale

La méthode d'inhibition latérale (equation 2.2) a été modifiée afin de tenir compte du chevauchement des filtres cochléaires dans le domaine spectral. Selon nos observations, cela permet d'obtenir une meilleure reconstruction du signal. L'équation suivante décrit la matrice d'inhibition latérale G utilisée par le FLCAA.

$$G_{i,j} =< |H_i(f)|.|H_j(f)| > \quad \text{pour i et j} = 1,..., \text{M} \tag{3.8}$$

Où $< . >$ est le produit scalaire, $|.|$ le module et $H_i(f)$ est la fonction de transfert normalisée des filtres (RoExp).

3.1.5 Critères d'optimisation du codage

Nous modifions l'équation 2.7 en intégrant l'utilisation des fenêtres à analyser($w[j]$). Ainsi, le codage tente de minimiser l'erreur quadratique moyenne entre chaque $w[j]$ au lieu du signal $s(t)$ en entier afin de permettre un traitement par fenêtre coulissante pour une implémentation temps réel. Pour le FLCAA, chaque fenêtre d'analyse est codée et évaluée indépendamment : il y a donc une fonction de coût pour chacun des $w[j]$. On présente à l'équation 3.9 la formule gouvernant l'optimisation du codage LCA.

$$E(t)[j] = \frac{1}{2} \|w[j] - \hat{w}[j]\|^2 + \lambda \sum_{m=1}^{M} C(a_m[j]) \tag{3.9}$$

λ correspond au paramètre de parcimonie, M au nombre de filtres cochléaires et C à la fonction de coût (équation 2.6). $\hat{w}[j]$ est la reconstruction de $w[j]$ présentée à l'équation 3.10. La minimisation de la fonction d'énergie est assurée par le mécanisme de seuillage. Le FLCAA utilise un seuillage dur.

3.1.6 Transformation de synthèse

Il est possible de reconstruire le signal suite à la transformation d'analyse avec fenêtre coulissante. Pour y arriver, on commence par la reconstruction indépendante de chaque fenêtre d'analyse, à l'aide des coefficients perceptuels parcimonieux $a_j[n]$ et de chaque filtres cochléaires ϕ_j, tel que décrit par l'équation suivante.

$$\hat{w}_i[n] = \sum_{j=1}^{M} \phi_{i,j} a_j[n] \tag{3.10}$$

La contribution de chaque o_j est déterminée, pour un indice de temps fixe i, à l'intérieur de la fenêtre d'analyse ($i = 1, ..., N$), avec M correspondant au nombre de filtres dans Φ. $\hat{w}_i[n]$ est un vecteur colonne de taille N et $a_j[n]$ est le $j^{ième}$ coefficient estimé à l'instant n. Il est à noter que chaque $\hat{w}_i$ peut être calculé dès que les coefficients $a_j[n]$ sont disponibles. La taille du tampon nécessaire pour une reconstruction temps réel doit être d'environ trois fois la taille de la fenêtre coulissante afin d'avoir toute l'information nécessaire. Ainsi, en ayant initialement suffisamment de coefficients en tampon, on peut reconstruire chaque fenêtre, en temps réel, lors de l'acquisition des coefficients de chaque fenêtre d'analyse.

De même, chaque $\hat{s}[n]$ peut être calculé dès que les N segements reconstruits $\hat{w}_i$ nécessaires sont disponibles pour l'échantillon n. Ainsi, en ayant initialement N segements reconstruits $\hat{w}_i$ en tampon, on peut reconstruire chaque échantillon n de $\hat{s}[n]$ subséquent, en temps réel, lors de l'acquisition de chaque $\hat{w}_i$.

$$\hat{s}[n] = \sum_{i=1}^{N} \hat{w}_i[n + 1 - i] \tag{3.11}$$

N correspond à la longueur de la réponse impulsionnelle du banc de filtres et, conséquemment, à la longueur de la fenêtre d'analyse. Pour reconstruire le signal $\hat{s}$ en entier on utilise $n = 1, 2, ..., L$.

L'équation de reconstruction 3.12 est obtenue en substituant l'équation 3.10 dans l'équation 3.11. Celle-ci correspond à une reconstruction causale. Cela signifie que l'information des coefficients précédents à l'instant $n + 1 - i$ est suffisante pour estimer s à l'instant n. La causalité de l'approche est favorable pour la reconstruction du signal en temps réel.

$$\hat{s}[n] = \sum_{i=1}^{N} \left(\sum_{j=1}^{M} \phi_{i,j} a_j[n + 1 - i] \right) \tag{3.12}$$

Validation de l'équation de reconstruction

Étant donné qu'il est possible d'interchanger les sommes dans l'équation 3.12 sans affecter le résultat de la reconstruction, il est possible d'obtenir l'équation équivalente suivante :

$$\hat{s}[n] = \sum_{j=1}^{M} \left(\sum_{i=1}^{N} o_{i,j} a_j[n + 1 - i] \right) \tag{3.13}$$

$a_j[n+1-i]$ est la séquence des coefficients associée à la base j. Pour un indice fixe j, $o_{i,j}$ est la réponse impulsionnelle de la base j. En calculant le résultat de la sommation de droite, on constate qu'elle correspond à une convolution, telle que décrite ci-dessous.

$$\sum_{i=1}^{N} \phi_{i,j} a_j[n+1-i] = \phi_j * a_j[n] \tag{3.14}$$

L'équation de reconstruction 3.15 est obtenue en substituant l'équation 3.14 dans l'équation 3.13.

$$\hat{s}[n] = \sum_{j=1}^{M} \phi_j * a_j[n] \tag{3.15}$$

Ainsi, la procédure de synthèse développée pour le FLCA est équivalente à utiliser un banc de filtres d'analyse/synthèse ayant des réponses impulsionnelles égales à ϕ_m.

CHAPITRE 4

Expériences et conditions expérimentales

4.1 Environnement et paramètres

On compare les représentations perceptuelles d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante à celles du FLCAA. Les deux techniques utilisent un banc de 128 filtres cochléaires et un algorithme de fenêtre coulissante. Leur différence est que le filtrage par banc de filtres cochléaires n'utilise pas le codage LCA tandis que le FLCAA l'utilise. Avec ce test, on tente de déterminer la validité du codage FLCAA et le potentiel du codage LCA avec le mécanisme de fenêtre coulissante. Également, la qualité des signaux reconstruits est comparée pour FLCAA et LCAASS (Pichevar *et al.* [29]). Notre implémentation du LCAASS présente deux différences fondamentales avec les travaux originaux de Pichevar *et al.*. Les filtres Gammatones ont étés remplacés par des filtres RoExp et la fonction de coût perceptuelle du LCAASS n'a pas été implémentée afin de faciliter la comparaison de la qualité des signaux reconstruits. On présume qu'il serait ultérieurement possible d'inclure la fonction de coût perceptuel dans notre FLCAA pour une meilleure reconstruction perceptive du signal.

Le filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante, le LCAASS utilisé dans ce mémoire et le FLCAA utilisent tous trois les filtres cochléaires de type RoExp (section 3.1.2). Deux configurations de 16 et 128 filtres sont implémentées pour le FLCAA et FLCAASS afin d'évaluer l'impact du changement de la taille du dictionnaire surcomplet sur le temps d'exécution et sur la qualité de la reconstruction. Le banc de 16 filtres est obtenu en enlevant 7 filtres sur 8 du banc de 128 filtres préalablement généré. Ainsi, on conserve une représentation de l'espace fréquentiel moins complète certe mais avec beaucoup moins de filtres ce qui diminue la durée nécessaire au codage. D'une part, le FLCAA utilise la matrice d'inhibition latérale $G_{i,j}$, décrite à l'équation 2.2, correspondante au niveau de similarité entre les réponses impulsionelles des filtres dans le domaine spectral. D'autre part, le LCAASS utilise la matrice d'inhibition latérale agrandie expliquée à la section 2.4 avec $q = 23$ (durée d'environ 1,5 ms), où q a été choisi comme valeur permettant un bon compromis entre la qualité de la reconstruction et la durée de codage. Lorsque le décalage entre les banc de filtres est plus petit, cela tend à augmenter la qualité de la reconstruction, la durée du codage et demande plus de mémoire.

Afin de poser le nombre d'itérations nécessaires au FLCAA, on valide expérimentalement l'effet de la variation du nombre d'itérations sur la qualité de la reconstruction du signal. On observe sur le tableau 5.1 que pour des signaux de musique qui sont plus complexes, le FLCAA est optimale vers 20 itérations. Cela démontre une convergence rapide du FLCAA pour atteindre une solution optimale. Ainsi, on pose le nombre d'itérations à 20 pour les expériences sur le FLCAA. On pose $\Delta/\tau = 0.37$ pour le FLCAA et $\Delta/\tau = 0.01$ pour le LCAASS car, on a déterminé expérimentalement, comme pour le nombre d'itérations, que ces valeurs sont optimales pour les signaux utilisées. Δ est le pas d'échantillonnage utilisé par l'algorithme pour discrétiser le signal et τ est le facteur d'inertie de l'algorithme. En d'autres mots, τ est le facteur déterminant la variabilité du codage entre deux itérations.

Le LCAASS quand à lui utilise un codage en 1000 itérations. Ce paramètre est posé de sorte à maximiser la qualité de la reconstruction tout en optimisant la durée du codage. Plus d'itérations rehausse la qualité de reconstruction mais le gain n'est pas significatif car la durée nécessaire au codage augmente de façon importante. Inversement, la réduction du nombre d'itérations réduit le temps d'exécution mais il y a une importante dégradation du signal. Cela, à cause de la taille du dictionnaire surcomplet du LCAASS qui nécessite plus d'itérations pour l'atteinte d'une solution optimale.

Toutes les simulations sont effectuées en utilisant un seuillage dur posé égal à $\lambda = 0.001$ car, un seuil dur permet d'atteindre une solution globalement optimale tel que décrit par les travaux originaux sur le LCA [32]. En plus de favoriser une bonne qualité de reconstruction, un seuil dur augmente la reproductibilité du codage en éliminant la variabilité des résultats causé par l'atteinte d'une solution localement optimale comme avec un seuil mou.

Les signaux sonores de test sont : la parole d'un homme disant /d/ /a/, une gamme jouée par un saxophone, le chant d'une femme, une chanson country folk et une chanson électronique. Ces signaux sonores ont été choisis pour couvrir un ensemble de sonorités permettant de démontrer la fonctionnalité et la polyvalence du FLCAA. Tous les signaux sonores, sauf la chanson électronique, proviennent d'une banque de sons standardisés et ont une fréquence d'échantillonnage de 32KHz. La chanson électronique à une fréquence d'échantillonnage de 44,1KHz. Les simulations sont effectuées sous Matlab R2012b sur un ordinateur desktop PC intel core 2 quad CPU Q6600 cadencé à 2.4GHz sur Windows 7 ultimate avec 4G de RAM.

4.2 Validation de la fonctionnalité et de la polyvalence du FLCAA

Cette expérience sert à valider que le FLCAA est fonctionnel et polyvalent en comparant les représentations perceptuelles du FLCAA avec celles des filtrages par banc de filtres cochléaires utilisant une fenêtre coulissante pour divers signaux sonores présentés à la section précédente. Si les représentations sont similaires, c'est-à-dire qu'on observe les mêmes structures sonores dans le temps à cause des distributions spectrales similaires, il est possible de valider l'implémentation du FLCAA. Cela est crucial pour entraîner correctement une méthode de composition musicale assistée par ordinateur à l'aide des sorties du FLCAA.

4.3 Qualité de la reconstruction

Cette expérience sert à déterminer le niveau de similarité entre le signal original et sa reconstruction pour les systèmes FLCAA et LCAASS. Une reconstruction fidèle améliore l'applicabilité pour tous systèmes ayant besoin de revenir dans le domaine original après analyse. Cela est particulièrement important pour les MCMAO recherchant obtenir une reconstruction de qualité. Les critères de performance retenus pour valider la qualité de la reconstruction sont le rapport signal sur bruit(RSB), le RSB segmentaire, la distorsion spectrale logarithmique(DSL) et l'évaluation perceptuelle de la qualité de la parole(EPQP).

4.4 Durée nécessaire au codage

Cette expérience sert à évaluer, pour les systèmes FLCAA et LCAASS, l'impact du changement de la taille du signal et du nombre de filtres cochléaires utilisés sur le temps nécessaire au codage. Les tests sont effectués avec le signal de parole /d/ /a/ dit par un homme.

Plus précisemment, nous modifions la taille du signal original (5.2) afin d'évaluer la capacité du système à traiter des signaux de plus en plus longs. Un système capable de traiter des signaux de longue durée est plus facilement intégrable dans une MCMAO que s'il est limité aux signaux de courte durée.

Également, nous modifions le nombre de filtres composant le dictionnaire surcomplet. Selon nos observations préliminaires, une analyse utilisant plus de filtres différents tend

à permettre une reconstruction plus fidèle car il y a moins de perte d'information. Ainsi, un système capable de fonctionner avec un plus grand nombre de filtres permettrait une meilleure qualité de reconstruction en général.

4.5 Robustesse

Cette expérience sert à déterminer si le FLCAA est robuste. Les tests sont effectués avec le signal de parole /d/ /a/ dit par un homme. La robustesse permet à un système de codage d'omettre le traitement d'une partie de l'information tout en obtenant une bonne reconstruction du signal. Cela est favorable, en télécommunication par exemple, lorsque l'information est possiblement corrompue durant l'acheminement ou l'acquisition. Aussi, un système démontrant une bonne robustesse peut accélérer le temps nécessaire au codage en omettant volontairement le traitement d'une partie de l'information.

La robustesse du FLCAA est testée en modifiant l, le nombre d'échantillons de décalage entre chaque fenêtre d'analyse (équation 3.1) et en comparant les signaux reconstruits entre eux. Ainsi, en augmentant l, le traitement d'une partie des fenêtres d'analyse est omis ce qui permet de simuler une perte d'information et ce qui a pour effet d'accélérer le temps de convergence car il y a moins de traitements nécessaires pour effectuer la reconstruction. Cette expérience n'est pas valide pour le LCAASS car la perte d'informations signifie un signal corrompu.

CHAPITRE 5

Résultats et discussions

5.1 Validation de la fonctionnalité et de la polyvalence du FLCAA

On présente aux figures 5.1, 5.2, 5.3, 5.4 et 5.5 les représentations perceptuelles obtenues à l'aide de filtrages par banc de filtres cochléaires utilisant une fenêtre coulissante et à l'aide de l'analyse FLCAA. On observe distinctement, dans tous les cas, une corrélation entre les représentations perceptuelles des deux techniques. En effet leurs fréquences fondamentales ainsi que leurs distributions spectrales concordent, ce qui tend à confirmer que la représentation FLCAA est sensée et pourrait servir à entraîner une méthode de composition musicale par ordinateur.

On observe que les représentations perceptuelles FLCAA contiennent plus d'information sur le signal analysé révélant une meilleure résolution spatiale et temporelle. Par exemple, sur la figure 5.2 on remarque dans la représentation perceptuelle FLCAA (à droite) du signal de parole /d/ /a/ qu'il y a plus de structures couvrant les filtres ayant des fréquences centrales basses (de 1 à 20) que pour la représentation perceptuelles sans codage LCA. Aussi, on remarque l'appartition de nouvelles informations entre les filtres 50 et 60 pour la durée entre $2x10^4$ et $2.15x10^4$ échantillons. De plus, vers $2x10^4$ échantillons la représentation FLCAA présente deux pointes et non une seule. On observe également que les contrastes sont forts en hautes fréquences car il y a amplification des caractéristiques sonores en raison de l'inhibition latérale et de l'élimination des coefficients trop faibles par seuillage. Vers $1{,}925x10^4$ échantillons on semble voir l'apparition du transitoire sur la représentation perceptuelle FLCAA mais cela peut être causé par des artefacts car les transitoires n'ont pas été observées sur les autres représentations perceptuelles analysées. Toutes ces informations sont potentiellement intéressantes pour l'entraînement d'une MC-MAO qui aura accès à plus de caractéristiques sonores pour établir son modèle. On observe sur les autres figures de cette section des résultats similaires quand à l'amélioration des résolutions fréquentielle et temporelle lors du codage FLCAA ce qui tend à démontrer la fonctionnalité et la polyvalence de celui-ci sur divers types de signaux. Ainsi, le choix du codage FLCAA en tant que technique d'analyse pour une méthode de composition musicale assistée par ordinateur est justifiable parce qu'il offre une haute résolution spatiale et

Comparaison des performances du FLCAA sur divers signaux en fonction du nombre d'itérations

Nb. itérations		Da Homme	Chant Femme	Electronique	Country folk	moyennes
5	RSB	18,72	23,93	11,88	12,80	16,83
	RSB seg.	16,61	23,43	12,38	12,81	16,31
	DSL	1,66	0,90	2,35	2,85	1,94
	Durée (s)	8,33	8,42	8,43	8,30	8,37
10	RSB	25,78	29,90	17,98	17,14	22,70
	RSB seg.	23,55	28,79	18,29	17,58	22,05
	DSL	1,03	0,69	2,36	2,58	1,67
	Durée (s)	13,43	13,48	13,30	13,48	13,42
20	RSB	26,38	29,10	20,03	18,82	23,58
	RSB seg.	25,42	28,29	20,20	20,32	23,56
	DSL	1,01	0,68	2,41	2,47	1,64
	Durée (s)	23,39	23,66	23,49	23,45	23,50
35	RSB	27,42	28,52	18,79	18,58	23,33
	RSB seg.	26,06	27,75	18,87	19,35	23,01
	DSL	0,97	0,67	2,53	2,44	1,65
	Durée (s)	38,89	38,91	38,27	38,58	38,66
50	RSB	27,93	28,40	17,50	17,80	22,91
	RSB seg.	25,66	27,57	17,57	18,19	22,25
	DSL	0,91	0,67	2,63	2,43	1,66
	Durée (s)	54,12	54,33	53,41	54,01	53,97
100	RSB	28,56	29,01	15,32	16,08	22,24
	RSB seg.	27,02	27,87	15,45	16,20	21,64
	DSL	0,89	0,65	2,83	2,48	1,71
	Durée (s)	104,17	104,87	104,29	104,12	104,36

Tableau 5.1 On présente les performances de reconstruction du FLCAA sur divers types de signaux, présentés à la fin de la section 4.1, en fonction du nombre d'itérations maximale alloué pour la convergence du codage. Le FLCAA utilise un banc de 128 filtres. La durée du codage (en seconde), le rapport signal sur bruit (RSB), le RSB segmentaire, la distortion spectrale logarithmique (DSL) sont présentés pour divers signaux sonores. On remarque que les performances du FLCAA s'améliore avec le nombre d'itérations pour les signaux simples (Da Homme et Chant Femme) mais qu'il est maximal rapidement pour les signaux complexes (Electronique et Country folk). De plus, ce tableau prouve la polyvalence du FLCAA en démontrant qu'on peut coder tous les types de signaux et obtenir une reconstruction de bonne qualité (rapport signal sur bruit d'environ 20dB ou plus) avec 10 à 20 itérations au maximum.

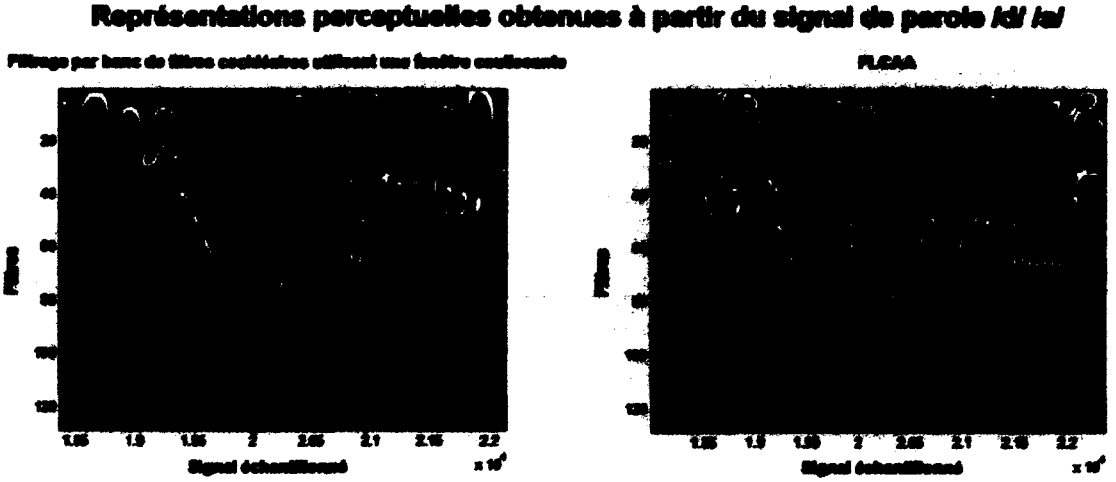

Figure 5.1 Comparaison entre les représentations perceptuelles du signal /d/ /a/ pour un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante et le FLCAA. Sur la gauche, on présente une section agrandie d'une représentation perceptuelle résultant d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante. Sur la droite, on présente une section agrandie de la représentation perceptuelle FLCAA. Les 2 images correspondent au même segment de signal. Les filtres en ordonnée sont représentés par leurs indices et positionnés selon leurs fréquences centrales. Plus l'indice est bas, plus la fréquence centrale du filtre est basse. On remarque que les deux images sont similaires et présentent la fréquence fondamentale du signal ainsi que la distribution spectrale du signal. Cela valide que la représentation perceptuelle du FLCAA a du sens et qu'elle pourra servir à entraîner une MCMAO. On dénote également qu'il y a beaucoup plus d'information présente dans la représentation perceptuelle FLCAA à cause du codage LCA qui rehausse et met en évidence les caractéristiques du signal. Spécifiquement, la résolution spatiale en basse fréquence, l'apparition d'information en moyenne fréquence et la définition de l'information en haute fréquence. Vers $1{,}925 \times 10^4$ échantillons on semble voir l'apparition du transitoire sur la représentation perceptuelle FLCAA mais cela peut être causé par des artefacts car les transitoires n'ont pas été observées sur les autres représentations perceptuelles analysées.

temporelle ce qui révèle plus d'information et ainsi augmente le potentiel de modélisation du signal. De plus, il est fonctionnel pour la musique et divers autres types de signaux sonores ce qui ne limite pas son utilisation par une MCMAO.

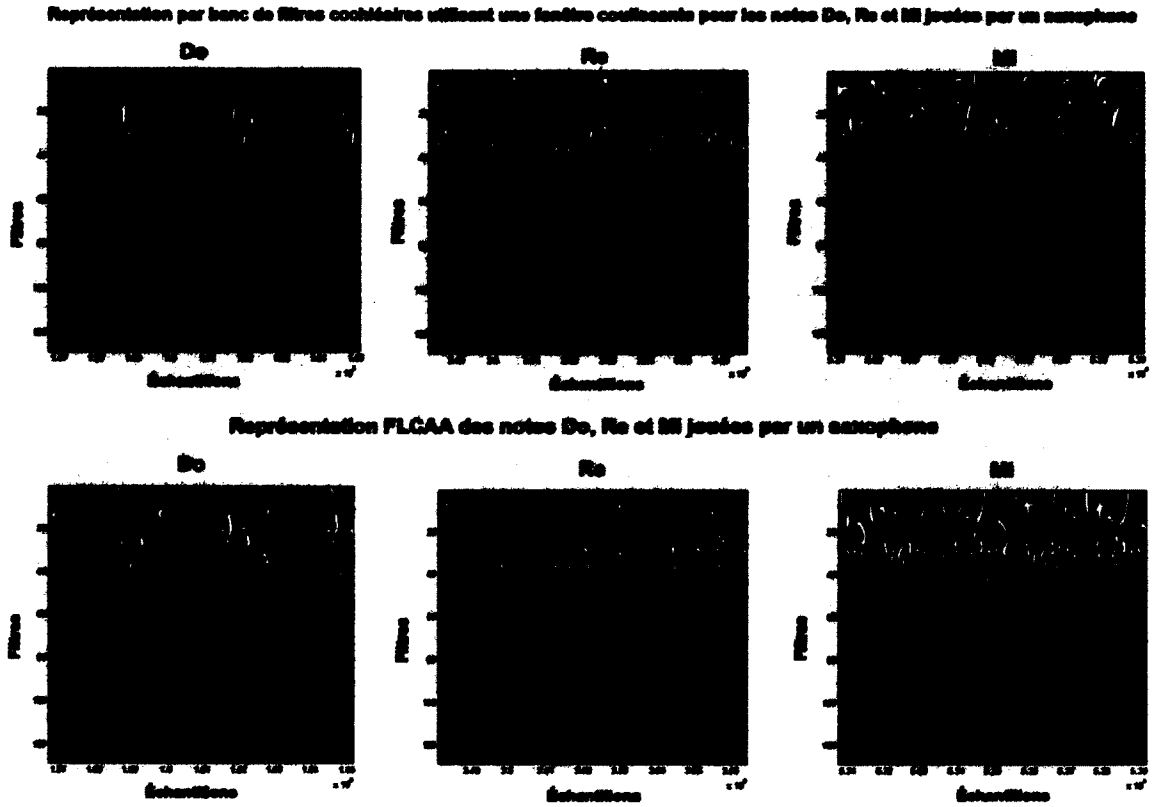

Figure 5.2 Comparaison entre les représentations perceptuelles d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante et le FLCAA pour les notes Do, Ré et Mi jouées par un saxophone. En haut, on présente les sections agrandies des représentations perceptuelles résultantes d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante pour les notes Do, Ré et Mi. En bas, on présente une section agrandie de la représentation perceptuelle FLCAA pour les notes Do, Ré et Mi. Les filtres en ordonnée sont représentés par leurs indices et positionnés selon leurs fréquences centrales. Plus l'indice est bas, plus la fréquence centrale du filtre est basse. On remarque que le FLCAA fait ressortir l'information du signal à cause de sa bonne résolution fréquencielle. Cela est observable par l'émergence de structures sonores autrement invisble tel qu'observé sur le Ré entre les filtres de 30 à 40 ainsi que sur le Mi pour le filtre 50. Les présents travaux n'évaluent toutefois pas la pertinence de ces nouvelles structures. On remarque également une meilleure résolution temporelle pour les représentations perceptuelles FLCAA.

5.2 Qualité de reconstruction

On observe au tableau 5.2 de meilleures reconstructions lorsque le banc de filtres est constitué de 128 filtres plutôt que de 16 filtres, et ce, autant pour le LCAASS que pour

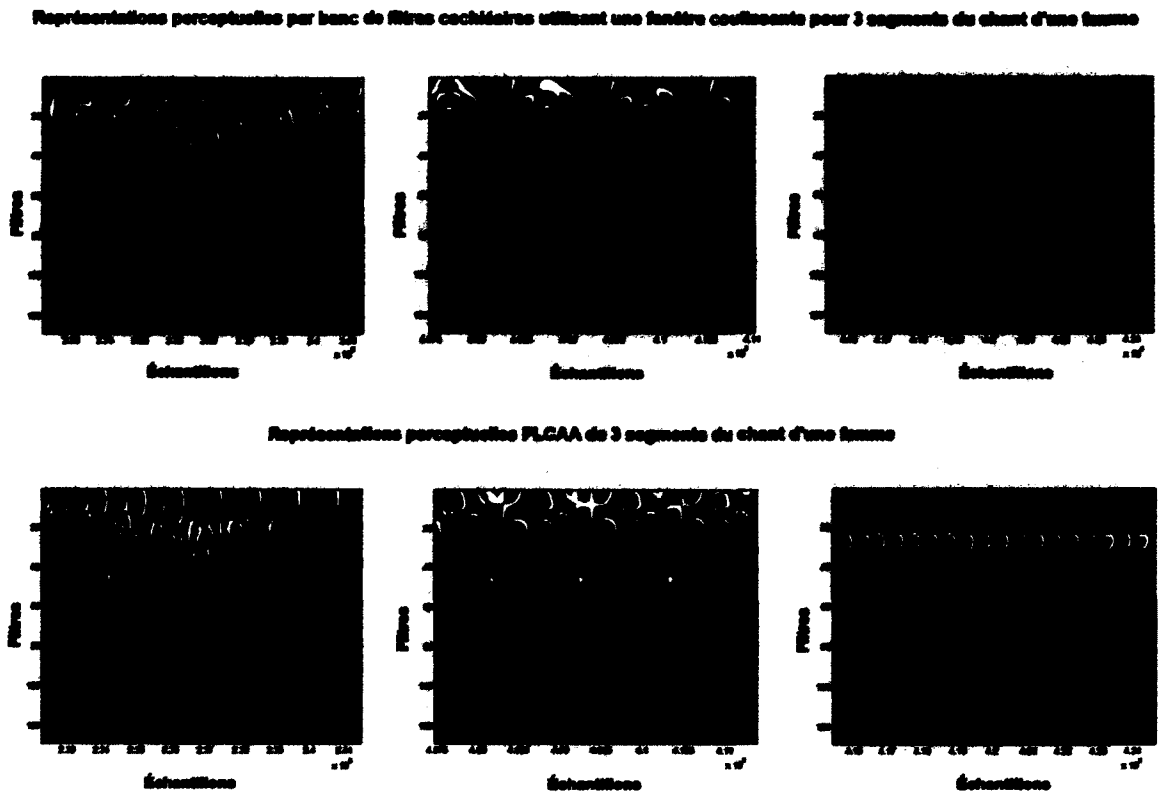

Figure 5.3 Comparaison entre les représentations perceptuelles d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante et le FLCAA pour le chant d'une femme. En haut, on présente les sections agrandies des représentations perceptuelles résultantes d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante pour trois segments différents de chant. En bas, on présente une section agrandie de la représentation perceptuelle FLCAA pour trois segments différents de chant. Les filtres en ordonnée sont représentés par leurs indices et positionnés selon leurs fréquences centrales. Plus l'indice est bas, plus la fréquence centrale du filtre est basse. On remarque que les trois représentations FLCAA présentent une meilleure résolution fréquencielle que l'autre technique. Cela est observable par les structures nouvelles des filtres 20 à 35 sur la représentation de droite et des filtres 40 à 50 pour les autres représentations FLCAA.

le FLCAA. Tel que mentionné précédemment, la reconstruction est meilleure lorsque les filtres représentent plus adéquatement l'espace à coder. Ainsi, il y a moins de perte d'information. Cela implique que la qualité de la reconstruction est directement liée au nombre de filtres et qu'ils favorisent l'obtention d'une reconstruction fidèle.

Comparaison des performances du LCAASS et du FLCAA

	LCAASS					
16 filtres	Nb Échant.	RSB	RSB seg.	DSL	Durée (s)	Durée (s)/Échant.
	516	16,36	16,36	1,31	0,76	0,0015
	1032	16,13	16,94	1,74	5,39	0,0052
	2064	15,95	15,11	1,77	20,57	0,0099
	4128	14,25	15,17	2,04	77,49	0,0188
128 filtres	Nb Échant.	RSB	RSB seg.	DSL	Durée (s)	Durée (s)/Échant.
	516	27,94	26,46	0,52	10,45	0,0202
	1032	24,99	24,7	0,84	40,76	0,0395
	2064	N/A	N/A	N/A	N/A	N/A
	4128	N/A	N/A	N/A	N/A	N/A
	FLCAA (20 itérations au maximum)					
16 filtres	Nb Échant.	RSB	RSB seg.	DSL	Durée (s)	Durée (s)/Échant.
	516	10,61	10,78	2,46	0,69	0,0013
	1032	13,58	13,4	2,3	1,62	0,0015
	2064	14,61	14,55	1,7	3,4	0,0016
	4128	15,33	15,24	1,67	7,04	0,0017
128 filtres	Nb Échant.	RSB	RSB seg.	DSL	Durée (s)	Durée (s)/Échant.
	516	12,8	13,05	1,73	1,26	0,0025
	1032	19,35	19,43	1,39	3,08	0,003
	2064	25,51	25,57	0,96	6,77	0,0033
	4128	25,64	25,72	0,9	13,96	0,0034

Tableau 5.2 Comparaison des performances de reconstruction du LCAASS (1000 itérations au maximum) et du FLCAA (20 itérations au maximum) pour le signal de parole /d//a/ dit par un homme. On présente les résultats des deux systèmes utilisant des bancs de filtres de 16 et 128 filtres et sections du signal de différentes durées. La durée du signal (en échantillons), le rapport signal sur bruit (RSB), le RSB segmentaire, la distortion spectrale logarithmique (DSL), la durée du codage (en secondes) et la durée de codage moyen (en secondes par échantillon) sont présentés. On ne peut réaliser l'évaluation perceptuelle de la qualité de la parole (EPQP) parce que les sons utilisés sont trop courts pour qu'il soit possible de calculer un indice EPQP significatif. N/A indique qu'il y a eu dépassement de la mémoire nécessaire pour réaliser le codage.

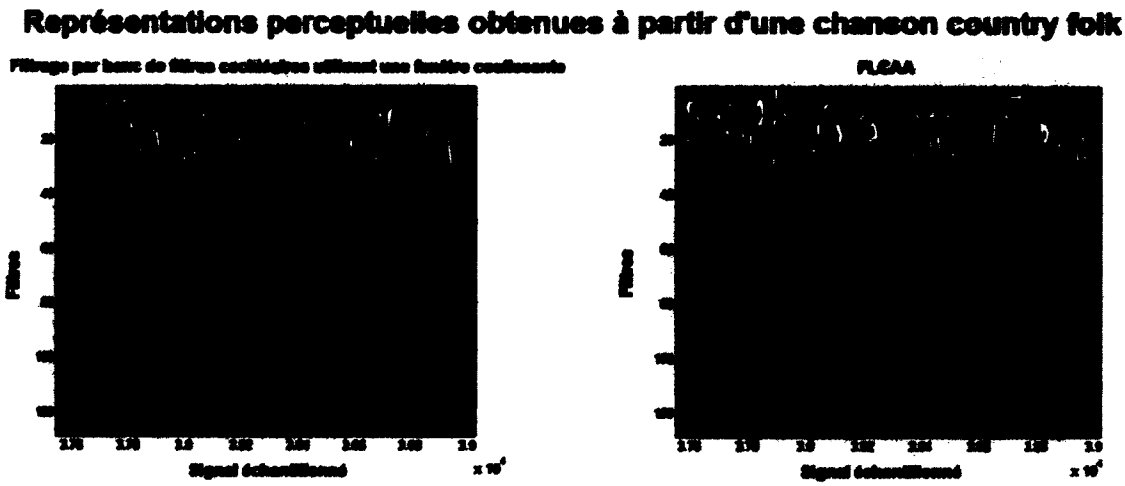

Figure 5.4 Comparaison entre les représentations perceptuelles d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante et le FLCAA pour un segment de chanson counrty folk. En haut, on présente les sections agrandies des représentations perceptuelles résultantes d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante pour un segment de chanson country folk. En bas, on présente une section agrandie de la représentation perceptuelle FLCAA pour un segment de chanson country folk. Les filtres en ordonnée sont représentés par leurs indices et positionnés selon leurs fréquences centrales. Plus l'indice est bas, plus la fréquence centrale du filtre est basse. On remarque une meilleure résolution temporelle sur la représentation FLCAA car les pics sont plus fins.

Dans toutes les simulations réalisées, le codage LCAASS permet une meilleure qualité de reconstruction du signal. Toutefois, il est possible d'optimiser le FLCAA afin de compenser la plus faible qualité de reconstruction de celui-ci en utilisant un plus grand nombre de filtres, ce qui n'affecte quasiment pas le temps de convergence (figure 5.6). Les expériences sur le gain potentiel maximal de qualité de reconstruction du FLCAA en augmentatant le nombre de filtres ne sont pas traités dans ce mémoire mais il serait pertinent d'évaluer ce gain dans des travaux futurs. On observe sur le tableau 5.2 la mention N/A qui indique que le système LCAASS manque de mémoire et est incapable de fonctionner lorsque le signal atteint une durée de 2064 échantillons et plus pour un banc de 128 filtres. Ainsi, on constate que la taille du signal à coder devient un facteur déterminant pour éviter la surcharge de la mémoire pour le LCAASS.

Également, on observe sur le tableau 5.1 que la qualité de la reconstrution FLCAA pour divers types de signaux sonores est de bonne qualité (RSB d'environ 20dB avec 10 à 20 itérations au maximum). Cela démontre la polyvalence du FLCAA.

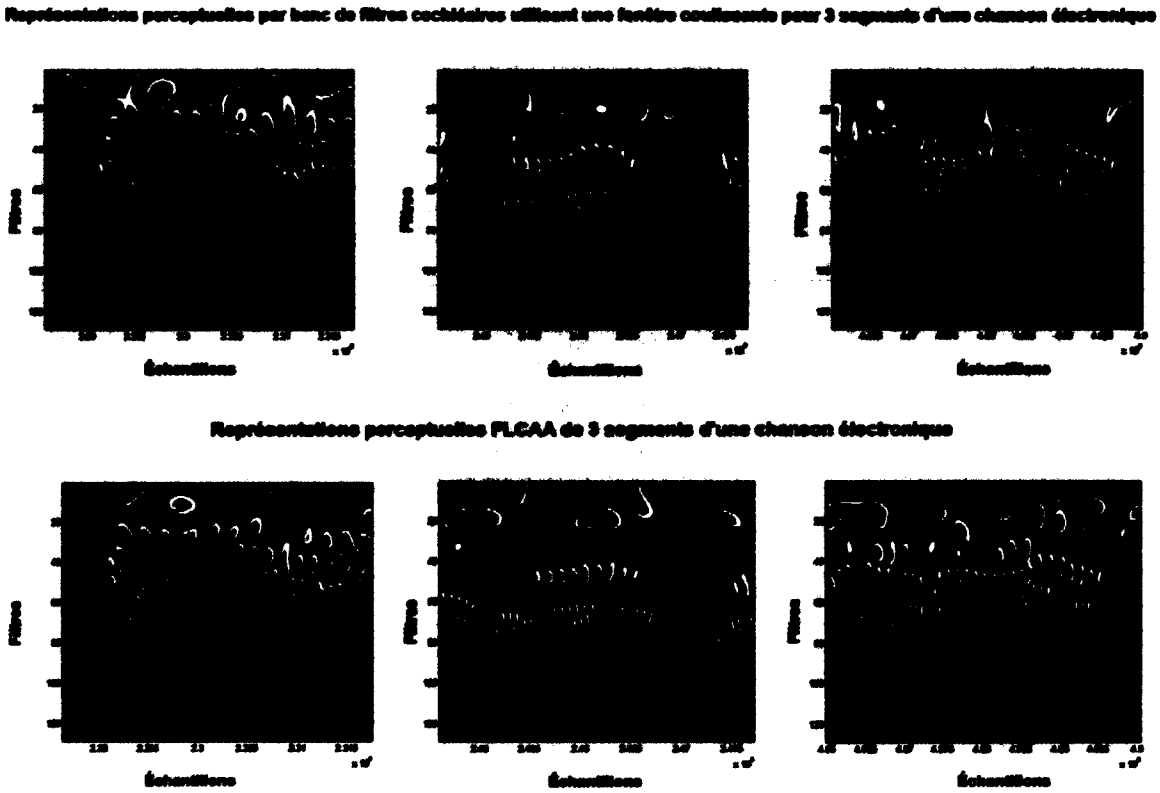

Figure 5.5 Comparaison entre les représentations perceptuelles d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante et le FLCAA pour trois segments de chanson electroniques. En haut, on présente les sections agrandies des représentations perceptuelles résultantes d'un filtrage par banc de filtres cochléaires utilisant une fenêtre coulissante pour trois segments de chanson electroniques. En bas, on présente une section agrandie de la représentation perceptuelle FLCAA pour trois segments de chanson electroniques. Les filtres en ordonnée sont représentés par leurs indices et positionnés selon leurs fréquences centrales. Plus l'indice est bas, plus la fréquence centrale du filtre est basse. On observe une bonne résolution spectrale car les structures sonores sont mises en évidence dans les représentations FLCAA.

5.3 Durée nécessaire au codage

La figure 5.6 présente la progression des durées nécessaires aux codages pour les deux techniques, en fonction de la taille du signal avec 16 et 128 filtres. Les durées nécessaires aux codages du LCAASS croissent de manière exponentielle en augmentant la durée du signal à coder. Cela est causé par l'intégration de l'information de phase dans le dictionnaire surcomplet (section 2.4). La croissance des durées nécessaires aux codages est encore plus importante lorsque le nombre de filtres augmente. À l'inverse, le temps de convergence du

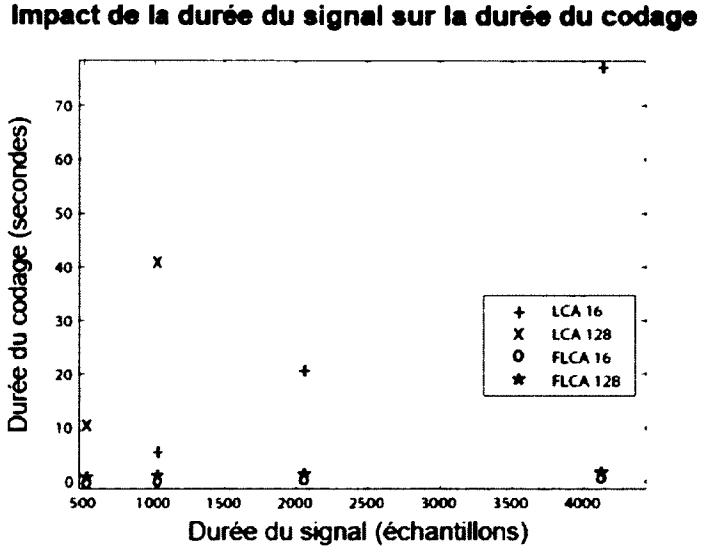

Figure 5.6 Progression des durées nécessaires aux codages (en secondes), du FLCAA et du LCAASS, en fonction de la taille du signal, exprimé en échantillons, avec 16 et 128 filtres. Les durées nécessaires aux codages pour le LCAASS croîssent de manière exponentielle selon le nombre d'échantillons à coder tandis que l'augmentation est linéaire pour le FLCAA. Également, l'impact de l'augmentation du nombre de filtres sur les durées nécessaires aux codages est quasi inexistant pour le FLCAA contrairement au LCAASS. En fait l'augmentation de la durée du codage FLCAA est tellement faible entre 500 et 4000 échantillons qu'on observe difficilement l'augmentation de la durée de codage sur cette figure. La légende différencie les implémentations avec 16 et 128 filtres.

FLCAA est très peu affecté par le nombre de filtres cochléaires ou par la durée du signal. On observe difficilement l'augmentation de la durée du codage entre 500 et 4000 échantillons en utilisant 16 ou 128 filtres tellement elle est faible. Cela démontre l'applicabilité supérieure de la méthode FLCAA pour des signaux moindrement longs.

De plus, comme le temps de convergence du FLCAA est peu affecté par l'augmentation du nombre de filtres cochléaires, il est possible d'utiliser un nombre élevé de filtres pour représenter le signal ce qui a un impact direct sur la qualité de la représentation.

5.4 Robustesse

La figure 5.7 présente l'impact de l'augmentation de l, le nombre d'échantillons entre deux fenêtres d'analyse successives, sur le EPQP, le RSB, le RSB segmentaire et le DSL sur la reconstruction du signal de parole « d » « a » dit par un homme. L'augmentation de l

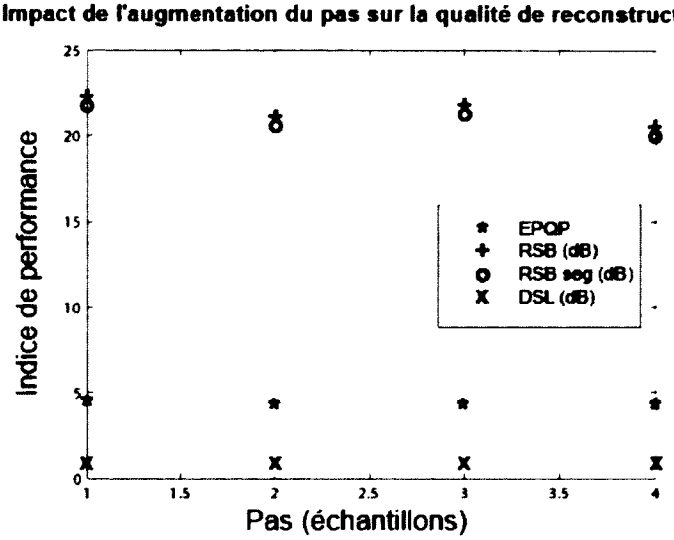

Figure 5.7 Évaluation de la robustesse du FLCAA en évaluant l'impact de l'omission de fenêtres d'analyse sur la durée du codage (en secondes) et sur la qualité de la reconstruction du FLCAA. On présente les résultats du codage du signal de parole /d/ /a/ dit par un homme avec un pas de $l = 1, 2, 3, 4$ échantillons entre les fenêtres d'analyse successives ce qui correspond à l'omission de 0,1,2 et 3 fenêtres d'analyse sur 4. On décale de plus en plus la distance entre deux fenêtre successives à analyser lorsque l croît, tel que décrit à la section 3.1.1. Ainsi, on observe une dégradation mineure du rapport signal sur bruit (RSB) et du RSB segmentaire plus l'on omet de fenêtres d'analyse. Toutefois, la distortion spectrale logarithmique (DSL) et l'évaluation perceptuelle de la qualité de la parole (EPQP) restent stables. Ces résultats démontrent la robustesse du FLCAA. La durée du codage, non présentée sur cette figure, est réduite linéairement selon le pourcentage de fenêtres d'analyse traitées en moins.

correspond à l'omission de $(l-1)$ fenêtres d'analyse sur l. Ainsi, en augmentant l, on décale de plus en plus la distance entre deux fenêtre successives, tel que décrit à la section 3.1.1. On observe que l'augmentation de l entraîne une dégradation mineure du rapport signal sur bruit (RSB) et du RSB segmentaire. Cependant, la distortion spectrale logarithmique (DSL) et l'évaluation perceptuelle de la qualité de la parole (EPQP) restent stables. Ces résultats démontrent la robustesse du FLCAA et indiquent qu'il est facile de diminuer la durée nécessaire au codage FLCAA en augmentant l sans trop affecter la qualité de la reconstruction. En fait, la durée du codage est réduite linéairement selon le pourcentage de fenêtres d'analyse traitées en moins. Toutefois, lorsque $l > 4$, on observe une dégradation importante du signal reconstruit ce qui signifie l'atteinte de la limite de robustesse à cause d'une perte d'information trop grande pour permettre une reconstruction fidèle du signal.

CHAPITRE 6

Conclusion

L'applicabilité du FLCAA a été démontrée selon diverses contraintes du monde réel. En effet, il permet de coder des signaux de toutes tailles, en temps réel, avec autant de filtres cochléaires que désiré, en plus de permettre une reconstruction robuste de bonne qualité sur divers types de signaux sonores.

Comme le FLCAA utilise une fenêtre coulissante pour encoder le signal, le dictionnaire surcomplet Φ n'a pas besoin d'inclure l'information de phase pour obtenir une bonne qualité de reconstruction. Cela signifie une durée de codage réduite et moins de mémoire nécessaire pour le codage FLCAA que pour le codage LCAASS, spécialement pour les signaux de longue durée. Notons que l'implémentation parallèle du FLCAA n'a pas été réalisée dans ce mémoire et que les durées de codage peuvent donc être optimisées davantage. Cela fera l'étude de travaux futures.

Le FLCAA permet une optimisation de la qualité de la reconstruction par l'amélioration du dictionnaire surcomplet servant à la représentation du signal. En effet, il est possible d'augmenter le nombre de filtres cochléaires pour obtenir une reconstruction plus fidèle du signal à coder sans avoir d'impact significatif sur le temps d'exécution. Cela pourrait permettre d'optimiser la reconstruction du codage FLCAA davantage en travaillant sur divers types de bancs de filtres inspirés de l'audition. Les expériences servant à tester le gain potentiel maximal du FLCAA en augmentant le nombre de filtres ou en modifiant la réponse impusionnelle des filtres ne sont pas traitées dans ce mémoire mais il serait pertinent de le faire lors de travaux futurs.

La robustesse du FLCAA contribue à son applicabilité car, il est possible d'omettre jusqu'à trois fenêtres d'analyse sur quatre sans que l'on observe une dégradation significative (figure 5.7). La robustesse du FLCAA peut être utilisée pour sécuriser la qualité de reconstruction ou pour accélérer l'exécution en réduisant le nombre de fenêtres à analyser.

On démontre également que notre méthode de reconstruction est équivalente à la convolution pour une analyse par banc de filtres. La différence majeure de notre approche est que la reconstruction peut être faite pièce par pièce à partir d'une représentation perceptuelle au lieu d'une convolution unique sur le signal complet (equation 3.15) comme pour le LCAASS. Cela signifie que l'on peut effectuer des reconstructions partielles (équa-

tion 3.10) qui seront intégrées dans le temps (équation 3.11) pour la reconstruction finale. Cette différence majeure dans notre méthode de reconstruction permet une implémentation aisée d'un FLCAA parallèle étant donné que chaque fenêtre d'analyse est codée et reconstruite indépendamment les unes des autres.

Par la qualité de sa représentation perceptuelle, par ses performances de codage et de reconstruction, par sa polyvalence et par son applicabilité, le FLCAA semble une technique appropriée pour entraîner diverses méthodes de composition musicale assistée par ordinateur. De plus, nous sommes convaincus que bien d'autres applications intégrant le FLCAA à des fins d'identification et de traitement de caractéristiques sonores verront le jour dans un futur rapproché.

LISTE DES RÉFÉRENCES

[1] Anders, T. (2003). *Composing Music by Composing Rules : Design and Usage of a Generic Music Constraint System.* PhD thesis, Queen's University, Belfast.

[2] Ando, D. et Iba, H. (2007). Interactive composition aid system by means of tree representation of musical phrase. Dans *IEEE Congress on Evolutionary Computation (September).* IEEE, Singapore, p. 4258–4265.

[3] Balavoine, A., Romberg, J. K. et Rozell, C. J. (2012). Convergence and rate analysis of neural networks for sparse approximation. *IEEE Trans. on Neural Networks and Learning Systems.*, volume 23, numéro 9, p. 1377–1389.

[4] Burton, A. R. et Vladimirova, T. (1999). Generation of musical sequences with genetic techniques. *Computer Music Journal*, volume 23, numéro 4, p. 59–73.

[5] Casey, M. (2005). Acoustic lexemes for organizing internet audio. *Contemporary Music Review*, volume 24, numéro 6, p. 489–508.

[6] Charles, A., Kressner, A. et Rozell, C. (2011). A causal locally competitive algorithm for the sparse decomposition of audio signals. Dans *Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE), 2011 IEEE DOI - 10.1109/DSP-SPE.2011.5739223.* p. 265–270.

[7] Chen, C.-C. J. et Miikkulainen, R. (2001). Creating melodies with evolving recurrent neural networks. Dans *Proceedings of the INNS-IEEE International Joint Conference on Neural Networks.* IEEE, Piscataway, NJ, p. 2241–2246.

[8] Chiu, S.-C. et Shan, M.-K. (2006). Computer music composition based on discovered music patterns. Dans *Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on.* volume 5. p. 4401–4406.

[9] Corrêa, D. C., Saito, J. H. et Abib, S. (2008). Composing music with bptt and lstm networks : Comparing learning and generalization aspects. Dans *Proceedings of the 2008 11th IEEE International Conference on Computational Science and Engineering - Workshops.* CSEWORKSHOPS '08. IEEE Computer Society, Washington, DC, USA, p. 95–100.

[10] Daugman, J. G. (1980). Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, volume 20, numéro 10, p. 847 – 856.

[11] Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A : Optics, Image Science, and Vision*, volume 2, numéro 7, p. 1160–1169.

[12] Davis, S. et Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, volume 28, numéro 4, p. 357–366.

[13] Dawkins, R. (1989). *The evolution of evolvability.* Addison-Wesley, Redwood City, CA, p. 201 - 220.

[14] Franklin, J. A. (2006). Recurrent neural networks for music computation. *INFORMS Journal on Computing*, volume 18, numéro 3, p. 321-338.

[15] Hiller, L. et Issacson, L. (1959). *Experimental music : composition with an electronic computer.* McGraw-Hill.

[16] Hoffman, M., Cook, P. et Blei, D. (2008). Data-driven recomposition using the hierarchical dirichlet process hidden markov model. *Proc. International Computer Music Conference.*

[17] Laden, B. et Keefe, D. (1989). *The Representation of Pitch in a Neural Network Model of Musical Chord Classification.* Technical report series, University of Washington, School of Music.

[18] Liang, P., Jordan, M. I. et Klein, D. (2009). Probabilistic grammars and hierarchical dirichlet processes. *The Oxford Handbook of Applied Bayesian Analysis.*

[19] Liu, Y. (1992). *Un détecteur perceptif de la hauteur tonale pour la parole téléphonique.* Mémoire de maîtrise, Université du Québec à Chicoutimi.

[20] Mallat, S. et Zhang, Z. (1993). Matching pursuits with time-frequency dictionaries. *Signal Processing, IEEE Transactions on*, volume 41, numéro 12, p. 3397 -3415.

[21] Mathews, M. V. (1963). The digital computer as a musical instrument. *Science*, volume 142, numéro 3592, p. 553-557.

[22] Mccormack, J. (1996). *Grammar Based Music Composition.* R. Stocker et. al., eds, ISO Press, Amsterdam, 320 - 336 p.

[23] Mermelstein, P. (1976). Distance measures for speech recognition-psychological and instrumental. Dans *Joint Workshop on Pattern Recognition and Artificial Intelligence.*

[24] Miranda, E. (2001). *Composing Music with Computers with Cdrom.* Music Technology Series, Focal Press.

[25] Moore, B. C. J. et Glasberg, B. R. (1983). Suggested formulae for calculating auditoryâfilter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, volume 74, numéro 3, p. 750-753.

[26] Mozer, M. C. (1994). Neural network music composition by prediction : exploring the benefits of psychoacoustic constraints and multi-scale processing. Dans *Connection Science.* p. 247-280.

[27] Oliwa, T. et Wagner, M. (2008). Composing music with neural networks and probabilistic finite-state machines. Dans *Applications of Evolutionary Computing.* Lecture Notes in Computer Science, volume 4974. Springer, p. 503-508.

[28] Patterson, R. D. (1976). Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America*, volume 59, numéro 3, p. 640-654.

[29] Pichevar, R., Najaf-Zadeh, H. et Mustière, F. (2010). Neural-based approach to perceptual sparse coding of audio signals. Dans *IEEE Joint Conference on Neural Networks*.

[30] Pichevar, R., Najaf-Zadeh, H., Thibault, L. et Lahdili, H. (2011). Auditory-inspired sparse representation of audio signals. *Speech Communication*, volume 53, numéro 5, p. 643 - 657.

[31] Rouat, J., Loiselle, S. et Molotchnikoff, S. (2011). Variable frame rate hierarchical analysis for robust speech recognition. Dans *Proc. of EEE/RSJ Int. Conf. on Intelligent Robots and Systems*.

[32] Rozell, C. J., Johnson, D. H., Baraniuk, R. G. et Olshausen, B. A. (2008). Sparse coding via thresholding and local competition in neural circuits. *Neural Computation*, volume 20, p. 2526 - 2563.

[33] Schwarz, D. (2004). *Data-Driven Concatenative Sound Synthesis*. Thèse de doctorat, Ircam - Centre Pompidou, Paris, France.

[34] Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica*, volume 4, p. 639–650.

[35] Sheikholharam, P. et Teshnehlab, M. (2008). Music composition using combination of genetic algorithms and kohonen grammar. Dans *Computational Intelligence and Design, 2008. ISCID '08. International Symposium on*. volume 1. p. 255–260.

[36] Skorokhod A. V., S. R. L. (1967). *Conditional Markov processes and their application to the theory of optimal control*. 184 - 187 p.

[37] Teh, Y. W., Jordan, M. I., Beal, M. J. et Blei, D. M. (2006). Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, volume 101, numéro 476, p. 1566–1581.

[38] Tropp, J. (2004). Greed is good : algorithmic results for sparse approximation. *Information Theory, IEEE Transactions on*, volume 50, numéro 10, p. 2231 - 2242.

[39] Vetterli, M. (1986). Filter banks allowing perfect reconstruction. *Signal Processing*, volume 10, numéro 3, p. 219–244.

[40] Wang, X. (2007). Neural coding strategies in auditory cortex. *Hearing Research*, volume 229, numéro 1 à 2, p. 81 - 93.

[41] Yang, X., Wang, K. et Shamma, S. (1992). Auditory representations of acoustic signals. *Information Theory, IEEE Transactions on*, volume 38, numéro 2, p. 824 839.